

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : <b>C12N 15/12, C07K 14/47, 16/18, G01N 33/566, C12Q 1/68, C12N 15/11, 15/62, A01K 67/027, A61K 38/00</b>		A2	(11) International Publication Number: <b>WO 00/58473</b>
			(43) International Publication Date: 5 October 2000 (05.10.00)
(21) International Application Number: PCT/US00/08621		(72) Inventors; and	
(22) International Filing Date: 31 March 2000 (31.03.00)		(75) Inventors/Applicants (for US only): SHIMKETS, Richard, A. [US/US]; 191 Leete Street, West Haven, CT 06516 (US). LEACH, Martin [GB/US]; 884 School Street, Webster, MA 01570 (US).	
(30) Priority Data:		(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).	
60/127,607	31 March 1999 (31.03.99) US		
60/127,636	2 April 1999 (02.04.99) US		
60/127,728	5 April 1999 (05.04.99) US		
09/540,763	30 March 2000 (30.03.00) US		
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
US	60/127,607 (CIP)		
Filed on	31 March 1999 (31.03.99)		
US	60/127,636 (CIP)		
Filed on	2 April 1999 (02.04.99)		
US	60/127,728 (CIP)		
Filed on	5 April 1999 (05.04.99)		
US	09/540,763 (CIP)		
Filed on	30 March 2000 (30.03.00)		
(71) Applicant (for all designated States except US): CURAGEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).		Published Without international search report and to be republished upon receipt of that report.	
(54) Title: NUCLEIC ACIDS INCLUDING OPEN READING FRAMES ENCODING POLYPEPTIDES; "ORFX"			
(57) Abstract			
<p>The present invention provides open reading frames ORFX, encoding isolated polypeptides, as well as polynucleotides encoding ORFX and antibodies that immunospecifically bind to ORFX or any derivative, variant, mutant, or fragment of the ORFX polypeptides, polynucleotides or antibodies. The invention additionally provides methods in which the ORFX polypeptide, polynucleotide and antibody are used in detection and treatment of a broad range of pathological states, as well as to other uses.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

5

## BACKGROUND OF THE INVENTION

The invention relates generally to nucleic acids and polypeptides encoded thereby, and methods of using these nucleic acids and polypeptides.

10

## SUMMARY OF THE INVENTION

The invention is based in part on the discovery of nucleic acids that include open reading frames encoding novel polypeptides, and on the polypeptides encoded thereby. The nucleic acids and polypeptides are collectively referred to herein as "ORFX".

15

Accordingly, in one aspect, the invention provides an isolated nucleic acid molecule (SEQ ID NO:2 $n$ -1, wherein  $n$  is an integer between 1-3161), that encodes novel polypeptide, or a fragment, homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid sequence encoding a polypeptide at least 85% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO:2 $n$ , wherein  $n$  is an integer between 1-3161. The nucleic acid can be, *e.g.*, a genomic DNA fragment, or a cDNA molecule.

20

Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a recombinant expression vector comprising any of the nucleic acid molecules described above.

25

In another aspect, the invention includes a pharmaceutical composition that includes an ORFX nucleic acid and a pharmaceutically acceptable carrier or diluent.



In a further aspect, the invention includes a substantially purified ORF polypeptide, *e.g.*, any of the ORFX polypeptides encoded by an ORFX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a ORFX polypeptide and a pharmaceutically acceptable carrier or diluent.

5 In a still a further aspect, the invention provides an antibody that binds specifically to an ORFX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including ORFX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a  
10 polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing an ORFX polypeptide by providing a cell containing a ORFX nucleic acid, *e.g.*, a vector that includes a ORFX nucleic  
15 acid, and culturing the cell under conditions sufficient to express the ORFX polypeptide encoded by the nucleic acid. The expressed ORFX polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous ORFX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying an ORFX polypeptide or nucleic  
20 acids in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a ORFX polypeptide by contacting ORFX polypeptide with a compound and determining whether the ORFX polypeptide activity is modified.

25 The invention is also directed to compounds that modulate ORFX polypeptide activity identified by contacting a ORFX polypeptide with the compound and determining whether the compound modifies activity of the ORFX polypeptide, binds to the ORFX polypeptide, or binds to a nucleic acid molecule encoding a ORFX polypeptide.

In a another aspect, the invention provides a method of determining the presence of or  
30 predisposition of an ORFX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of ORFX polypeptide in the subject sample.

The amount of ORFX polypeptide in the subject sample is then compared to the amount of ORFX polypeptide in a control sample. An alteration in the amount of ORFX polypeptide in the subject protein sample relative to the amount of ORFX polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the ORFX is detected using a ORFX antibody.

In a further aspect, the invention provides a method of determining the presence of or predisposition of an ORFX-associated disorder in a subject. The method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the ORFX nucleic acid in the subject nucleic acid sample. The amount of ORFX nucleic acid sample in the subject nucleic acid is then compared to the amount of an ORFX nucleic acid in a control sample. An alteration in the amount of ORFX nucleic acid in the sample relative to the amount of ORFX in the control sample indicates the subject has a tissue proliferation-associated disorder.

In a still further aspect, the invention provides method of treating or preventing or delaying a ORFX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a ORFX nucleic acid, a ORFX polypeptide, or an ORFX antibody in an amount sufficient to treat, prevent, or delay a tissue proliferation-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

## DETAILED DESCRIPTION OF THE INVENTION

The invention provides novel polypeptides and nucleotides encoded thereby. The polynucleotides and their encoded polypeptides can be grouped according to the functions played by their gene products. Such functions include, structural proteins, proteins from which associated with metabolic pathways fatty acid metabolism, glycolysis, intermediary metabolism, calcium metabolism, proteases, and amino acid metabolism, etc.

Included in the invention are 3161 novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to as "ORFX nucleic acids" or ORFX polynucleotides" and the corresponding encoded polypeptide is referred to as a "ORFX polypeptide" or ORFX protein". For example, an ORFX nucleic acid according to the invention is a nucleic acid including an ORF1 nucleic acid, and an ORF polypeptide according to the invention is a polypeptide that includes the amino acid sequence of an ORF1 polypeptide.

Unless indicated otherwise, "ORFX" is meant to refer to any of the ORF1-3161 sequences disclosed herein.

Table 1 provides a summary of the ORFX nucleic acids and their encoded polypeptides are summarized in Table 1. Nucleic acid sequences and polypeptide sequences for ORFX nucleic acids according to the invention is provided in the section of the specification entitled "Disclosed Sequences of ORFX Nucleic Acid and Polypeptide Sequences."

Column 1 of Table 1, entitled "ORF #", denotes an ORF number assigned to a nucleic acid containing an open reading frame according to the invention.

Column 2 of Table 1, entitled "Internal Identification number (Nucleic Acid Sequence Identification Number, Polypeptide Sequence Identification Number), provides an internal identification number for the indicated ORF, along with sequence identification numbers (SEQ ID NOs.) corresponding to the indicated ORF. In general, for an ORF $n$  according to the invention (wherein  $n$  is any integer from 1 to 3161), a nucleic acid corresponding to the ORF is SEQ ID NO:2 $n$ -1, and an amino acid sequence encoded by the ORF is SEQ ID NO:2 $n$ . For example, a nucleic acid sequence corresponding to an ORF1 nucleic acid is SEQ ID NO:1, and a polypeptide sequence corresponding to an ORF1 polypeptide is SEQ ID NO:2. Similarly, a

nucleic acid sequence corresponding to an ORF4 nucleic acid is SEQ ID NO:7, and a polypeptide sequence corresponding to an ORF4 polypeptide is SEQ ID NO:8; a nucleic acid sequence corresponding to an ORF198 nucleic acid sequence is SEQ ID NO:395, and a polypeptide sequence corresponding to an ORF198 polypeptide is SEQ ID NO:396. Nucleic acid sequences and polypeptide sequences for ORFX nucleic acids according to the invention are provided in the section of the specification entitled "Disclosed Sequences of ORFX Nucleic Acid and Polypeptide Sequences."

Column 2 of Table 1, entitled "Protein Similarity", lists previously described proteins that are related to polypeptides encoded by the ORFs. Genbank identifiers for the previously described proteins are provided. These can be retrieved from <http://www.ncbi.nlm.nih.gov/>.

To determine similarity to previously described proteins, polypeptides encoded by ORFX DNA sequences were tested using the Framesearch Algorithm against a nonredundant version of the GenPept Database from NCBI/Genbank. DNA sequences that had a score of '90' or above (Framesearch algorithm score, Edelman et. al. GCG Genetics) to a known protein were selected. Open reading frames were extended beyond the region of the protein matched using standard DNA translation and codon tables. Novel proteins that lacked a protein match were translated against the standard genetic codons and proteins with an ORF at least 80 amino acids and containing a Methionine start are included in the Table.

Column 3 of Table 3, entitled "Protein Domains", lists previously described protein domains, designated by pfam entries, that are present in polypeptides encoded by the ORFs. Also included in column 3 are proteins in which these domains are present. The pfam entries can be retrieved from <http://pfam.wustl.edu/>. DNA sequences were translated in all six frames and tested using the Hmmer Algorithm against the Pfam Database (References to the algorithm and Pfam database can be found at <http://pfam.wustl.edu>). Translated DNA sequences that matched a protein domain entry in the Pfam database AND had a score of 7.5' were selected.

Column 4 of Table 3, entitled "Protein Classification", lists the type of classification assigned for the protein, based on its homology. Examples of proteins in the classification include the following proteins:

### **Amylases**

Amylase is responsible for endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides. Variations in amylase gene may be indicative of delayed maturation and of various amylase producing neoplasms and carcinomas.

### 5       **Amyloid**

The serum amyloid A (SAA) proteins comprise a family of vertebrate proteins that associate predominantly with high density lipoproteins (HDL). The synthesis of certain members of the family is greatly increased in inflammation. Prolonged elevation of plasma SAA levels, as in chronic inflammation, 15 results in a pathological condition, called amyloidosis, which affects the liver, kidney and spleen and which is characterized by the highly insoluble accumulation of SAA in these tissues. Amyloid selectively inhibits insulin-stimulated glucose utilization and glycogen deposition in muscle, while not affecting adipocyte glucose metabolism. Deposition of fibrillar amyloid proteins intraneuronally, as neurofibrillary tangles, extracellularly, as plaques and in blood vessels, is characteristic of both Alzheimer's disease and aged Down's syndrome. 15 Amyloid deposition is also associated with type II diabetes mellitus.

### **Angiopoeitin**

Members of the angiopoeitin/fibrinogen family have been shown to stimulate the generation of new blood vessels, inhibit the generation of new blood vessels, and perform several roles in blood clotting. This generation of new blood vessels, called angiogenesis, is also an 20 essential step in tumor growth in order for the tumor to get the blood supply it needs to expand. Variation in these genes may be predictive of any form of heart disease, numerous blood clotting disorders, stroke, hypertension and predisposition to tumor formation and metastasis. In particular, these variants may be predictive of the response to various antihypertensive drugs and chemotherapeutic and anti-tumor agents.

### 25       **Apoptosis-related proteins**

Active cell suicide (apoptosis) is induced by events such as growth factor withdrawal and toxins. It is controlled by regulators, which have either an inhibitory effect on programmed cell

death (anti-apoptotic) or block the protective effect of inhibitors (pro-apoptotic). Many viruses have found a way of countering defensive apoptosis by encoding their own anti-apoptosis genes preventing their target-cells from dying too soon. Variants of apoptosis related genes may be useful in formulation of anti-aging drugs.

5           **Cadherin, Cyclin, Polymerase, Oncogenes, Histones, Kinases**

Members of the cell division/cell cycle pathways such as cyclins, many transcription factors and kinases, DNA polymerases, histones, helicases and other oncogenes play a critical role in carcinogenesis where the uncontrolled proliferation of cells leads to tumor formation and eventually metastasis. Variation in these genes may be predictive of predisposition to any form  
10 of cancer, from increased risk of tumor formation to increased rate of metastasis. In particular, these variants may be predictive of the response to various chemotherapeutic and anti-tumor agents.

**Colony-stimulating factor-related proteins**

Granulocyte/macrophage colony-stimulating factors are cytokines that act in  
15 hematopoiesis by controlling the production, differentiation, and function of 2 related white cell populations of the blood, the granulocytes and the monocytes-macrophages.

**Complement-related proteins**

Complement proteins are immune associated cytotoxic agents, acting in a chain reaction to exterminate target cells to that were opsonized (primed) with antibodies, by forming a  
20 membrane attack complex (MAC). The mechanism of killing is by opening pores in the target cell membrane. Variations in 20 complement genes or their inhibitors are associated with many autoimmune disorders. Modified serum levels of complement products cause edemas of various tissues, lupus (SLE), vasculitis, glomerulonephritis, renal failure, hemolytic anemia, thrombocytopenia, and arthritis. They interfere with mechanisms of ADCC (antibody dependent  
25 cell cytotoxicity), severely impair immune competence and reduce phagocytic ability. Variants of complement genes may also be indicative of type I diabetes mellitus, meningitis neurological disorders such as nemaline myopathy, neonatal hypotonia, muscular disorders such as congenital myopathy and other diseases.

### **Cytochrome**

The respiratory chain is a key biochemical pathway which is essential to all aerobic cells. There are five different cytochromes involved in the chain. These are heme bound proteins which serve as electron carriers. Modifications in these genes may be predictive of ataxia  
5 areflexia, dementia and myopathic and neuropathic changes in muscles. Also, association with various types of solid tumors.

### **Kinesins**

Kinesins are tubulin molecular motors that function to transport organelles within cells and to move chromosomes along microtubules during cell division. Modifications of these genes  
10 may be indicative of neurological disorders such as Pick disease of the brain, tuberous sclerosis.

### **Cytokines, Interferon, Interleukin**

Members of the cytokine families are known for their potent ability to stimulate cell growth and division even at low concentrations. Cytokines such as erythropoietin are cell-specific in their growth stimulation; erythropoietin is useful for the stimulation of the  
15 proliferation of erythroblasts. Variants in cytokines may be predictive for a wide variety of diseases, including cancer predisposition.

### **G-protein coupled receptors**

G-protein coupled receptors (also called R7G) are an extensive group of hormones, neurotransmitters, odorants and light receptors which transduce extracellular signals by  
20 interaction with guanine nucleotide-binding (G) proteins. Alterations in genes coding for G-coupled proteins may be involved in and indicative of a vast number of physiological conditions. These include blood pressure regulation, renal dysfunctions, male infertility, dopamine associated cognitive, emotional, and endocrine functions, hypercalcemia, chondrodysplasia and osteoporosis, pseudohypoparathyroidism, growth retardation and  
25 dwarfism.

**Thioesterases**

Eukaryotic thiol proteases are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. Variants of thioester associated genes may be predictive of neuronal disorders and mental illnesses such as Ceroid Lipoffiscinosis, Neuronal 1, Infantile, Santavuori disease and more.

The key to the molecule type is as follows:

	<b>Abbrev:</b>	<b>Title:</b>
10	amylase	amylase protein
	amylaseinhib	amylase inhibitor
	amyloid	amyloid protein
15	apoptosis	apoptosis associated protein
	apoptosisinhib	apoptosis inhibitors
	apoptosisrecep	apoptosis receptors
	ATPase_associated	ATPase associated protein
	biotindep	biotin dependent enzyme/protein
20	cadherin	cadherin protein
	calcium_channel	calcium channel protein
	carboxylase	carboxylase protein
	cathepsin	cathepsin/carboxypeptidases
	cathepsininhib	cathepsin/carboxypeptidase inhibitor
25	chloride_channel	chloride channel protein
	collagen	collagen
	complement	complement protein
	complementrecept	complement receptor protein
	complementinhib	complement inhibitor
30	csf	colony stimulating factor
	csfrecept	colony stimulating factor receptor
	cyclin	cyclin protein
	cyto450	cytochrome p450 protein
	cytochrome	cytochrome related protein
35	deaminase	deaminase
	dehydrogenase	dehydrogenase
	desaturase	desaturase
	dna_rna_bind	DNA/RNA binding protein/factor
	dna_rna_inhib	DNA/RNA binding protein/factor inhibitor
40	dynein	dynein



	elastase	elastase
	elastaseinhib	elastase inhibitor
	eph	EPH family of tyrosine kinases
	esterase	esterase
5	esteraseinhib	esterase inhibitor
	fgf	fibroblast growth factor
	fgfreceptor	fibroblast growth factor receptor
	gaba	GABA receptor
10	glucoamylase	glucoamylase
	glucoronidase	glucoronidase
	glycoprotein	glycoprotein
	Guanylyl	guanylylate cyclase
	helicase	helicase
	histone	histone
15	HOM	homologous
	homeobox	homeobox protein
	hydrolase	hydrolase
	hydroxysteroid	hydroxysteroid associated protein
	hypoxanthine	hypoxanthine associated protein
20	immunoglob	immunoglobulin
	immunoglobrecept	immunoglobulin receptor
	interferon	interferon
	interleukin	interleukin
	interleukinrecept	interleukin receptor
25	isomerase	isomerase
	isomeraseinhibitor	isomerase inhibitor
	isomerasereceptor	isomerase receptor
	kinase	kinase
30	kinaseinhibitor	kinase inhibitor
	kinasereceptor	kinase receptor
	kinesin	kinesin
	laminin	laminin associated protein
	lipase	lipase
	metallothionein	metallothionein
35	MHC	major histocompatibility complex
	misc_channel	miscellaneous channel
	ngf	nerve growth factor
	nuci_recpt	nuclear receptor
	nuclease	nuclease
40	oncogene	oncogene associated protein
	oxidase	oxidase
	oxygenase	oxygenase
	peptidase	peptidase
	peroxidase	peroxidase
45	phosphatase	phosphatase
	phosphataseinhib	phosphatase inhibitor

	phosphorylase PIR	phosphorylase PIR DATABASE (release 56, 29-OCT-1998)
5	polymerase potassium_channel prostaglandin protease proteaseinhib reductase	polymerase potassium channel protein prostaglandin protease protease inhibitor reductase
10	ribosomalprot RTR	ribosomal associated protein EMBLDATABASE translated entries not to be incorporated into SWISS-PROT (20-JUL-1998)
15	SIM SPTR	similar EMBL DATABASE translated entries to be incorporated into SWISS-PROT (20-JUL-1998)
20	struct sulfotransferase SWP	structural associated protein sulfotransferase SWISS-PROT DATABASE (release 18-OCT-1998)
25	SWPN synthase tgf tgfreceptor thioesterase thiolase tm7	SWISS-PROT Update (release 11-NOV-98) synthase transforming growth factor transforming growth factor receptor thioesterase thiolase
30	tnf traffic tnfreceptor TRN	seven transmembrane domain G-protein coupled receptor necrosis factor receptor tumor necrosis factor tumor trafficking associated protein EMBL DATABASE translated entries update (20-JUL-1998)
35	transcriptfactor transferase transport tubulin ubiquitin unclassified	transcription factor transferase transport protein tubulin ubiquitin
40	water channel	Protein not categorized into one of the aforementioned protein families water channel protein

Column 5 of Table 1, entitled, "Cells or Tissues in Which Gene is Expressed", denotes tissues, represented by five digit numbers, in which RNA homologous to the ORF nucleic acid sequences is present. Tissues or cells corresponding to the numbers are provided in Table 2.

ORFX nucleic acids, and their encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, various ORFX nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families indicated in Table 1, and/or according to the presence of domains and sequence relatedness to previously described proteins as summarized in Table 1.

ORFX nucleic acids and polypeptides according to the invention can also be used to identify cell types listed in Table 1 for an indicated ORFX according to the invention. Additional utilities for ORFX nucleic acids and polypeptides according to the invention are disclosed herein.

#### ORFX Nucleic Acids

The novel nucleic acids of the invention include those that encode an ORFX or ORFX-like protein, or biologically active portions thereof. The nucleic acids include nucleic acids encoding polypeptides that include the amino acid sequence of one or more of SEQ ID NO:2*n*, wherein *n* = 1 to 3161. The encoded polypeptides can thus include, *e.g.*, the amino acid sequences of SEQ ID NO: 2, 4, 6, 8, 10, . . . , 6310, 6312, 6314, 6316, 6318, 6320, and/or 6322.

In some embodiments, a nucleic acid encoding a polypeptide having the amino acid sequence of one or more of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) includes the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a fragment thereof. Additionally, the invention includes mutant or variant nucleic acids of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a fragment thereof, any of whose bases may be changed from the disclosed sequence while still encoding a protein that maintains its ORFX-like activities and physiological functions. The invention further includes the complement of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), including fragments, derivatives,

analogues and homolog thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

Also included are nucleic acid fragments sufficient for use as hybridization probes to identify ORFX-encoding nucleic acids (*e.g.*, ORFX mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of ORFX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated ORFX nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a complement of any of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n*=1 to 3161) as a hybridization probe, ORFX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to ORFX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a complement thereof. Oligonucleotides may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in

is one that is sufficiently complementary to the nucleotide sequence shown in of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161),, thereby forming a stable duplex.

- 5           As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term “binding” means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect.
- 10       Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

- Moreover, the nucleic acid molecule of the invention can comprise only a portion of the
- 15       nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of ORFX. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively,
- 20       and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but
- 25       differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

- Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or
- 30       analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the

invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489, which is incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of ORFX polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a ORFX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human ORFX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) as well as a polypeptide having ORFX activity. Biological activities of the ORFX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human ORFX polypeptide.

The nucleotide sequence determined from the cloning of the human ORFX gene allows for the generation of probes and primers designed for use in identifying the cell types disclosed and/or cloning ORFX homologues in other cell types, *e.g.*, from other tissues, as well as ORFX homologues from other mammals. The probe/primer typically comprises a substantially purified

oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161); or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161);  
5 or of a naturally occurring mutant of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161).

Probes based on the human ORFX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be  
10 used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a ORFX protein, such as by measuring a level of a ORFX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting ORFX mRNA levels or determining whether a genomic ORFX gene has been mutated or deleted.

"A polypeptide having a biologically active portion of ORFX" refers to polypeptides  
15 exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of ORFX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), that encodes a polypeptide having a ORFX biological activity (biological activities of the ORFX  
20 proteins are summarized in Table 1), expressing the encoded portion of ORFX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of ORFX. For example, a nucleic acid fragment encoding a biologically active portion of ORFX can optionally include a domain as shown in Table 1, column 4.

#### 25 ORFX variants

The invention further encompasses nucleic acid molecules that differ from the disclosed ORFX nucleotide sequences due to degeneracy of the genetic code. These nucleic acids thus encode the same ORFX protein as that encoded by the nucleotide sequence shown in SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). In another embodiment, an isolated nucleic acid molecule of  
30 the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161).



In addition to the human ORFX nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of ORFX may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the ORFX gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a ORFX protein, preferably a mammalian ORFX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the ORFX gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in ORFX that are the result of natural allelic variation and that do not alter the functional activity of ORFX are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding ORFX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the ORFX cDNAs of the invention can be isolated based on their homology to the human ORFX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

In another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding ORFX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at  $T_m$ , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of any of SEQ ID NO:2*n*-1 (wherein  $n = 1$  to 3161) corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein  $n = 1$  to 3161), or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are

hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and  
5 Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A  
10 non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at  
15 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981, *Proc Natl Acad Sci USA* 78: 6789-6792.

#### Conservative mutations

In addition to naturally-occurring allelic variants of the ORFX sequence that may exist in  
20 the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), thereby leading to changes in the amino acid sequence of the encoded ORFX protein, without altering the functional ability of the ORFX protein. For example, nucleotide substitutions  
25 leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of ORFX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the ORFX proteins of the present  
30 invention, are predicted to be particularly unamenable to alteration.

Amino acid residues that are conserved among members of an ORFX family members are predicted to be less amenable to alteration. For example, an ORFX protein according to the present invention can contain at least one domain (*e.g.*, as shown in Table 1) that is a typically conserved region in an ORFX family member. As such, these conserved domains are not likely to be amenable to mutation. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved among members of the ORFX family) may not be as essential for activity and thus are more likely to be amenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding ORFX proteins that contain changes in amino acid residues that are not essential for activity. Such ORFX proteins differ in amino acid sequence from any of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161), yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161), more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO:2.

An isolated nucleic acid molecule encoding a ORFX protein homologous to the protein of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) can be created by introducing one or more nucleotide substitutions, additions or deletions into the corresponding nucleotide sequence, *i.e.* SEQ ID NO:2*n*-1 for the corresponding *n*, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline,

phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in ORFX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a ORFX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for ORFX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant ORFX protein can be assayed for (1) the ability to form protein:protein interactions with other ORFX proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant ORFX protein and a ORFX receptor; (3) the ability of a mutant ORFX protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g.*, avidin proteins); (4) the ability to bind BRA protein; or (5) the ability to specifically bind an anti-ORFX protein antibody.

#### **Antisense**

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire ORFX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a ORFX protein of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) or antisense nucleic acids complementary to a ORFX nucleic acid sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding ORFX. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid

residues (*e.g.*, the protein coding region of a human ORFX that corresponds to any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161)). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding ORFX. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding ORFX disclosed herein (*e.g.*, SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161)), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of ORFX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of ORFX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of ORFX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (*v*), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (*v*), 5-methyl-2-thiouracil,

3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a ORFX protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

#### Ribozymes and PNA moieties

Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are

carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme.

- 5 Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave ORFX mRNA transcripts to thereby inhibit translation of ORFX mRNA. A ribozyme having specificity for a ORFX-encoding
- 10 nucleic acid can be designed based upon the nucleotide sequence of a ORFX DNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1 (wherein  $n = 1$  to 3161)). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a ORFX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742.
- 15 Alternatively, ORFX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

- Alternatively, ORFX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the ORFX (*e.g.*, the ORFX promoter and/or
- 20 enhancers) to form triple helical structures that prevent transcription of the ORFX gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

- In various embodiments, the nucleic acids of ORFX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of
- 25 the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has
- 30 been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide



synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of ORFX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of ORFX can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of ORFX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of ORFX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or

the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

### ORFX polypeptides

The novel protein of the invention includes the ORFX-like protein whose sequence is provided in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in FIG. 1 while still encoding a protein that maintains its ORFX-like activities and physiological functions, or a functional fragment thereof. For example, the invention includes the polypeptides encoded by the variant ORFX nucleic acids described above. In the mutant or variant protein, up to 20% or more of the residues may be so changed.

In general, an ORFX-like variant that preserves ORFX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above. Furthermore, without limiting the scope of the invention, positions of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) may be substituted such that a mutant or variant protein may include one or more substitutions

The invention also includes isolated ORFX proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-ORFX antibodies. In one embodiment, native ORFX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, ORFX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a ORFX

protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the ORFX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of ORFX protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of ORFX protein having less than about 30% (by dry weight) of non-ORFX protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-ORFX protein, still more preferably less than about 10% of non-ORFX protein, and most preferably less than about 5% non-ORFX protein. When the ORFX protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of ORFX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of ORFX protein having less than about 30% (by dry weight) of chemical precursors or non-ORFX chemicals, more preferably less than about 20% chemical precursors or non-ORFX chemicals, still more preferably less than about 10% chemical precursors or non-ORFX chemicals, and most preferably less than about 5% chemical precursors or non-ORFX chemicals.

Biologically active portions of a ORFX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the ORFX protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:2 that include fewer amino acids than the full length ORFX proteins, and exhibit at least one activity of a ORFX protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the ORFX protein. A biologically active portion of a ORFX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a ORFX protein of the present invention may contain at least one of the above-identified domains conserved between the FGF family of proteins. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native ORFX protein.

In an embodiment, the ORFX protein has an amino acid sequence shown in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). In other embodiments, the ORFX protein is substantially homologous to any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) and retains the functional activity of the protein of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161), yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below.

Accordingly, in another embodiment, the ORFX protein is a protein that comprises an amino acid sequence at least about 45% homologous, and more preferably about 55, 65, 70, 75, 80, 85, 90, 95, 98 or even 99% homologous to the amino acid sequence of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) and retains the functional activity of the ORFX proteins of the corresponding polypeptide having the sequence of SEQ ID NO:2*n* (wherein *n* = 1 to 3161).

#### Determining homology between two or more sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See, *Needleman and Wunsch* 1970 *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a

degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO:2n-1 (wherein n = 1 to 3161).

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of positive residues.

#### **Chimeric and fusion proteins**

The invention also provides ORFX chimeric or fusion proteins. As used herein, a ORFX "chimeric protein" or "fusion protein" includes a ORFX polypeptide operatively linked to a non-ORFX polypeptide. A "ORFX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to ORFX, whereas a "non-ORFX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the ORFX protein, *e.g.*, a protein that is different from the ORFX protein and that is derived from the same or a different organism. Within a ORFX fusion protein the ORFX polypeptide can correspond to all or a portion of a ORFX protein. In one embodiment, a ORFX fusion protein comprises at least one biologically active portion of a ORFX protein. In another embodiment, a ORFX fusion protein comprises at least two biologically active portions of a

ORFX protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the ORFX polypeptide and the non-ORFX polypeptide are fused in-frame to each other. The non-ORFX polypeptide can be fused to the N-terminus or C-terminus of the ORFX polypeptide.

5 For example, in one embodiment a ORFX fusion protein comprises a ORFX polypeptide operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds that modulate ORFX activity (such assays are described in detail below).

10 In another embodiment, the fusion protein is a GST-ORFX fusion protein in which the ORFX sequences are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant ORFX.

15 In yet another embodiment, the fusion protein is a ORFX protein containing a heterologous signal sequence at its N-terminus. For example, the native ORFX signal sequence can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of ORFX can be increased through use of a heterologous signal sequence.

20 In another embodiment, the fusion protein is a ORFX-immunoglobulin fusion protein in which the ORFX sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The ORFX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ORFX ligand and a ORFX protein on the surface of a cell, to thereby suppress ORFX-mediated signal transduction *in vivo*. In one nonlimiting example, a contemplated ORFX ligand of the invention is an ORFX receptor. The ORFX-immunoglobulin fusion proteins can be used to modulate the bioavailability of a ORFX cognate ligand. Inhibition of the ORFX ligand/ORFX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (e.g., promoting or inhibiting) cell survival. Moreover, the ORFX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-ORFX antibodies in a subject, to purify ORFX ligands, and in screening assays to identify molecules that inhibit the interaction of  
30 ORFX with a ORFX ligand.

A ORFX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A ORFX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the ORFX protein.

#### **ORFX agonists and antagonists**

The present invention also pertains to variants of the ORFX proteins that function as either ORFX agonists (mimetics) or as ORFX antagonists. Variants of the ORFX protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the ORFX protein. An agonist of the ORFX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the ORFX protein. An antagonist of the ORFX protein can inhibit one or more of the activities of the naturally occurring form of the ORFX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the ORFX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the ORFX proteins.

Variants of the ORFX protein that function as either ORFX agonists (mimetics) or as ORFX antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the ORFX protein for ORFX protein agonist or antagonist activity. In one

embodiment, a variegated library of ORFX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of ORFX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential ORFX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of ORFX sequences therein. There are a variety of methods which can be used to produce libraries of potential ORFX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential ORFX sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477.

#### Polypeptide libraries

In addition, libraries of fragments of the ORFX protein coding sequence can be used to generate a variegated population of ORFX fragments for screening and subsequent selection of variants of a ORFX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a ORFX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the ORFX protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of ORFX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors,



transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify ORFX variants (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).

#### Anti-ORFX Antibodies

The invention further encompasses antibodies and antibody fragments, such as  $F_{ab}$  or  $(F_{ab})_2$ , that bind immunospecifically to any of the proteins of the invention.

An isolated ORFX protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind ORFX using standard techniques for polyclonal and monoclonal antibody preparation. Full-length ORFX protein can be used. Alternatively, the invention provides antigenic peptide fragments of ORFX for use as immunogens. The antigenic peptide of ORFX comprises at least 4 amino acid residues of the amino acid sequence shown in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). The antigenic peptide encompasses an epitope of ORFX such that an antibody raised against the peptide forms a specific immune complex with ORFX. The antigenic peptide may comprise at least 6 aa residues, at least 8 aa residues, at least 10 aa residues, at least 15 aa residues, at least 20 aa residues, or at least 30 aa residues. In one embodiment of the invention, the antigenic peptide comprises a polypeptide comprising at least 6 contiguous amino acids of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161).

In an embodiment of the invention, epitopes encompassed by the antigenic peptide are regions of ORFX that are located on the surface of the protein, *e.g.*, hydrophilic regions. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each incorporated herein by reference in their entirety.

As disclosed herein, an ORFX protein sequence of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161), or derivatives, fragments, analogs or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind these protein components. The term "antibody" as used herein refers to immunoglobulin molecules and

immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen, such as ORFX. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain,  $F_{ab}$  and  $F_{(ab)_2}$  fragments, and an  $F_{ab}$  expression library. In a specific embodiment, antibodies to human ORFX proteins are disclosed. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to a ORFX protein sequence of any of SEQ ID NO:2n (wherein n = 1 to 3161) or derivative, fragment, analog or homolog thereof. Some of these proteins are discussed below.

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, recombinantly expressed ORFX protein or a chemically synthesized ORFX polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against ORFX can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of ORFX. A monoclonal antibody composition thus typically displays a single binding affinity for a particular ORFX protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular ORFX protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (see Kohler & Milstein, 1975 *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (see Kozbor, *et al.*, 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND

CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, *et al.*, 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Each of the above citations are incorporated herein by reference in their entirety

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to a ORFX protein (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F<sub>ab</sub> expression libraries (see *e.g.*, Huse, *et al.*, 1989 *Science* 246: 1275-1281) to allow rapid and effective identification of monoclonal F<sub>ab</sub> fragments with the desired specificity for a ORFX protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. See *e.g.*, U.S. Patent No. 5,225,539. Each of the above citations are incorporated herein by reference. Antibody fragments that contain the idiotypes to a ORFX protein may be produced by techniques known in the art including, but not limited to: (i) an F<sub>(ab)<sub>2</sub></sub> fragment produced by pepsin digestion of an antibody molecule; (ii) an F<sub>ab</sub> fragment generated by reducing the disulfide bridges of an F<sub>(ab)<sub>2</sub></sub> fragment; (iii) an F<sub>ab</sub> fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F<sub>v</sub> fragments.

Additionally, recombinant anti-ORFX antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 86/01533; U.S. Pat. No. 4,816,567; European Patent Application No. 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *PNAS* 84:3439-3443; Liu *et al.* (1987) *J Immunol.* 139:3521-3526; Sun *et al.* (1987) *PNAS* 84:214-218; Nishimura *et al.* (1987) *Cancer Res* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; Shaw *et al.* (1988), *J. Natl Cancer Inst* 80:1553-1559; Morrison (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; U.S. Pat. No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525;

Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J Immunol* 141:4053-4060.

Each of the above citations are incorporated herein by reference.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of a ORFX protein is facilitated by generation of hybridomas that bind to the fragment of a ORFX protein possessing such a domain. Antibodies that are specific for one or more domains within a ORFX protein, *e.g.*, the domain spanning the first fifty amino-terminal residues specific to ORFX when compared to FGF-9, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-ORFX antibodies may be used in methods known within the art relating to the localization and/or quantitation of a ORFX protein (*e.g.*, for use in measuring levels of the ORFX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for ORFX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds [hereinafter "Therapeutics"].

An anti-ORFX antibody (*e.g.*, monoclonal antibody) can be used to isolate ORFX by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-ORFX antibody can facilitate the purification of natural ORFX from cells and of recombinantly produced ORFX expressed in host cells. Moreover, an anti-ORFX antibody can be used to detect ORFX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the ORFX protein. Anti-ORFX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or

phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

#### ORFX Recombinant Vectors and Host Cells

5 Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding ORFX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can  
10 be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are  
15 replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the  
20 invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the  
25 recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation  
30 system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements

(*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, ORFX proteins, mutant forms of ORFX, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of ORFX in prokaryotic or eukaryotic cells. For example, ORFX can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (1) to increase expression of recombinant protein; (2) to increase the solubility of the recombinant protein; and (3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

5 One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons  
10 for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the ORFX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*,  
15 (1987) *EMBO J* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).

Alternatively, ORFX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*,  
20 SF9 cells) include the pAc series (Smith *et al.* (1983) *Mol Cell Biol* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J*  
25 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells. See, *e.g.*, Chapters 16 and 17 of Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory,  
30 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv Immunol* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to ORFX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews--Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant



host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, ORFX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding ORFX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) ORFX protein. Accordingly, the invention further provides methods for producing ORFX protein using the host cells of the invention. In one embodiment,

the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding ORFX has been introduced) in a suitable medium such that ORFX protein is produced. In another embodiment, the method further comprises isolating ORFX from the medium or the host cell.

## 5            **Transgenic animals**

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which ORFX-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous ORFX sequences  
10 have been introduced into their genome or homologous recombinant animals in which endogenous ORFX sequences have been altered. Such animals are useful for studying the function and/or activity of ORFX and for identifying and/or evaluating modulators of ORFX activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal  
15 includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous  
20 recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous ORFX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing ORFX-encoding  
25 nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The human ORFX DNA sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human ORFX gene, such as a mouse ORFX gene, can be isolated based on hybridization to  
30 the human ORFX cDNA (described further above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of

expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the ORFX transgene to direct expression of ORFX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan 1986, In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the ORFX transgene in its genome and/or expression of ORFX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding ORFX can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a ORFX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the ORFX gene. The ORFX gene can be a human gene (*e.g.*, SEQ ID NO:2n-1 (wherein  $n = 1$  to 3161)), but more preferably, is a non-human homologue of a human ORFX gene. For example, a mouse homologue of human ORFX gene of SEQ ID NO:2n-1 (wherein  $n = 1$  to 3161) can be used to construct a homologous recombination vector suitable for altering an endogenous ORFX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous ORFX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous ORFX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous ORFX protein). In the homologous recombination vector, the altered portion of the ORFX gene is flanked at its 5' and 3' ends by additional nucleic acid of the ORFX gene to allow for homologous recombination to occur between the exogenous ORFX gene carried by the vector and an endogenous ORFX gene in an embryonic stem cell. The additional flanking ORFX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector. See *e.g.*, Thomas *et al.* (1987) *Cell* 51:503 for a description of

homologous recombination vectors. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced ORFX gene has homologously recombined with the endogenous ORFX gene are selected (see e.g., Li *et al.* (1992) *Cell* 69:915).

The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras. See e.g., Bradley 1987, In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Curr Opin Biotechnol* 2:823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, e.g., Lakso *et al.* (1992) *PNAS* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.* (1997) *Nature* 385:810-813. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of

this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

### Pharmaceutical Compositions

The ORFX nucleic acid molecules, ORFX proteins, and anti-ORFX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a ORFX protein or anti-ORFX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use

as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by any of a number of routes, *e.g.*, as described in U.S. Patent Nos. 5,703,055. Delivery can thus also include, *e.g.*, intravenous injection, local administration (see U.S. Pat. No. 5,328,470) or stereotactic injection (see *e.g.*, Chen *et al.* (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

#### **Additional Uses and Methods of the Invention**

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: (a) screening assays; (b) detection assays (*e.g.*, chromosomal mapping, cell and tissue typing, forensic biology), (c) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and (d) methods of treatment (*e.g.*, therapeutic and prophylactic).

The isolated nucleic acid molecules of the invention can be used to express ORFX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect ORFX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a ORFX gene, and to modulate ORFX activity, as described further below. In addition, the ORFX proteins can be used to screen drugs or compounds that modulate the ORFX activity or expression as well as to treat disorders characterized by insufficient or excessive production of ORFX protein, for



example proliferative or differentiative disorders, or production of ORFX protein forms that have decreased or aberrant activity compared to ORFX wild type protein. In addition, the anti-ORFX antibodies of the invention can be used to detect and isolate ORFX proteins and modulate ORFX activity.

5 This invention further pertains to novel agents identified by the above described screening assays and uses thereof for treatments as described herein.

### Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, 10 peptidomimetics, small molecules or other drugs) that bind to ORFX proteins or have a stimulatory or inhibitory effect on, for example, ORFX expression or ORFX activity.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a ORFX protein or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained 15 using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are 20 applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc Natl Acad Sci U.S.A.* 90:6909; Erb *et al.* (1994) *Proc Natl Acad Sci U.S.A.* 91:11422; Zuckermann *et al.* (1994) *J Med Chem* 37:2678; Cho *et al.* (1993) 25 *Science* 261:1303; Carrell *et al.* (1994) *Angew Chem Int Ed Engl* 33:2059; Carell *et al.* (1994) *Angew Chem Int Ed Engl* 33:2061; and Gallop *et al.* (1994) *J Med Chem* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), on chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner U.S. Pat. No. 5,223,409), spores (Ladner USP '409), 30 plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and

Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc Natl Acad Sci U.S.A.* 87:6378-6382; Felici (1991) *J Mol Biol* 222:301-310; Ladner above.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a ORFX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the ORFX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the ORFX protein or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the test compound to preferentially bind to ORFX or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the ORFX protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of ORFX or a biologically active portion thereof can be accomplished, for example, by determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule. As used herein, a "target molecule" is a molecule with which a ORFX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a ORFX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule

associated with the internal surface of a cell membrane or a cytoplasmic molecule. A ORFX target molecule can be a non-ORFX molecule or a ORFX protein or polypeptide of the present invention. In one embodiment, a ORFX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.*, a signal generated by binding of a compound to a membrane-bound ORFX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with ORFX.

Determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a ORFX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a ORFX protein or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the ORFX protein or biologically active portion thereof. Binding of the test compound to the ORFX protein can be determined either directly or indirectly as described above. In one embodiment, the assay comprises contacting the ORFX protein or biologically active portion thereof with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the test compound to preferentially bind to ORFX or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting ORFX protein or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the ORFX protein or

biologically active portion thereof. Determining the ability of the test compound to modulate the activity of ORFX can be accomplished, for example, by determining the ability of the ORFX protein to bind to a ORFX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of ORFX can be accomplished by determining the ability of the ORFX protein further modulate a ORFX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting the ORFX protein or biologically active portion thereof with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the ORFX protein to preferentially bind to or modulate the activity of a ORFX target molecule.

The cell-free assays of the present invention are amenable to use of both the soluble form or the membrane-bound form of ORFX. In the case of cell-free assays comprising the membrane-bound form of ORFX, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of ORFX is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton<sup>®</sup> X-100, Triton<sup>®</sup> X-114, Thesit<sup>®</sup>, Isotridecypoly(ethylene glycol ether)<sub>n</sub>, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl)dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either ORFX or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to ORFX, or interaction of ORFX with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For

example, GST-ORFX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or ORFX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of ORFX binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either ORFX or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated ORFX or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with ORFX or target molecules, but which do not interfere with binding of the ORFX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or ORFX trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the ORFX or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the ORFX or target molecule.

In another embodiment, modulators of ORFX expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of ORFX mRNA or protein in the cell is determined. The level of expression of ORFX mRNA or protein in the presence of the candidate compound is compared to the level of expression of ORFX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of ORFX expression based on this comparison. For example, when expression of ORFX mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of ORFX mRNA or protein expression. Alternatively, when expression of ORFX mRNA or protein is less (statistically significantly less) in the presence of the candidate

compound than in its absence, the candidate compound is identified as an inhibitor of ORFX mRNA or protein expression. The level of ORFX mRNA or protein expression in the cells can be determined by methods described herein for detecting ORFX mRNA or protein.

In yet another aspect of the invention, the ORFX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, *e.g.*, U.S. Pat. No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J Biol Chem* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins that bind to or interact with ORFX ("ORFX-binding proteins" or "ORFX-bp") and modulate ORFX activity. Such ORFX-binding proteins are also likely to be involved in the propagation of signals by the ORFX proteins as, for example, upstream or downstream elements of the ORFX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for ORFX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a ORFX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with ORFX.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

#### Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a

minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample.

The ORFX sequences of the present invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Pat. No. 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the ORFX sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The ORFX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), as described above, can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

### **Predictive Medicine**

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining ORFX protein and/or nucleic acid expression as well as ORFX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant ORFX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with ORFX protein, nucleic acid expression or activity. For example, mutations in a ORFX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with ORFX protein, nucleic acid expression or activity.

Another aspect of the invention provides methods for determining ORFX protein, nucleic acid expression or ORFX activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX in clinical trials.

### **Use of Partial ORFX Sequences in Forensic Biology**

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen



found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, that can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NOs: \_\_ are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the ORFX sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of one or more of SEQ ID NO:2*n*-1 (where *n* = 1 to 3161), having a length of at least 20 bases, preferably at least 30 bases.

The ORFX sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or label-able probes that can be used, for example, in an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue, etc. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such ORFX probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, ORFX primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

### **Predictive Medicine**

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining ORFX protein and/or nucleic acid expression as well as ORFX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant ORFX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with ORFX protein, nucleic

acid expression or activity. For example, mutations in a ORFX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with ORFX protein, nucleic acid expression or activity.

5 Another aspect of the invention provides methods for determining ORFX protein, nucleic acid expression or ORFX activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

10 Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX in clinical trials.

These and other agents are described in further detail in the following sections.

#### Diagnostic Assays

15 Other conditions in which proliferation of cells plays a role include tumors, restenosis, psoriasis, Dupuytren's contracture, diabetic complications, Kaposi's sarcoma and rheumatoid arthritis.

An ORFX polypeptide may be used to identify an interacting polypeptide a sample or tissue. The method comprises contacting the sample or tissue with ORFX, allowing formation of a complex between the ORFX polypeptide and the interacting polypeptide, and detecting the complex, if present.

20 The proteins of the invention may be used to stimulate production of antibodies specifically binding the proteins. Such antibodies may be used in immunodiagnostic procedures to detect the occurrence of the protein in a sample. The proteins of the invention may be used to stimulate cell growth and cell proliferation in conditions in which such growth would be favorable. An example would be to counteract toxic side effects of chemotherapeutic agents on, for example, hematopoiesis and platelet formation, linings of the gastrointestinal tract, and hair follicles. They may also be used to stimulate new cell growth in neurological disorders including, for example, Alzheimer's disease. Alternatively, antagonistic treatments may be administered in which an antibody specifically binding the ORFX-like proteins of the invention

30

would abrogate the specific growth-inducing effects of the proteins. Such antibodies may be useful, for example, in the treatment of proliferative disorders including various tumors and benign hyperplasias.

Polynucleotides or oligonucleotides corresponding to any one portion of the ORFX nucleic acids of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) may be used to detect DNA containing a corresponding ORF gene, or detect the expression of a corresponding ORFX gene, or ORFX-like gene. For example, an ORFX nucleic acid expressed in a particular cell or tissue, as noted in Table 2, can be used to identify the presence of that particular cell type.

An exemplary method for detecting the presence or absence of ORFX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting ORFX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes ORFX protein such that the presence of ORFX is detected in the biological sample. An agent for detecting ORFX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to ORFX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length ORFX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to ORFX mRNA or genomic DNA, as described above. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting ORFX protein is an antibody capable of binding to ORFX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect ORFX mRNA, protein, or genomic DNA in a biological sample *in vitro* as

well as *in vivo*. For example, *in vitro* techniques for detection of ORFX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of ORFX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of ORFX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of ORFX protein include introducing into a subject a labeled anti-ORFX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting ORFX protein, mRNA, or genomic DNA, such that the presence of ORFX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of ORFX protein, mRNA or genomic DNA in the control sample with the presence of ORFX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of ORFX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting ORFX protein or mRNA in a biological sample; means for determining the amount of ORFX in the sample; and means for comparing the amount of ORFX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect ORFX protein or nucleic acid.

#### **Prognostic Assays**

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant ORFX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with ORFX protein, nucleic acid expression or activity in, *e.g.*, proliferative or differentiative disorders such as hyperplasias, tumors, restenosis, psoriasis, Dupuytren's

contracture, diabetic complications, or rheumatoid arthritis, etc.; and glia-associated disorders such as cerebral lesions, diabetic neuropathies, cerebral edema, senile dementia, Alzheimer's disease, etc. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant ORFX expression or activity in which a test sample is obtained from a subject and ORFX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of ORFX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant ORFX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant ORFX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder, such as a proliferative disorder, differentiative disorder, glia-associated disorders, etc. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant ORFX expression or activity in which a test sample is obtained and ORFX protein or nucleic acid is detected (*e.g.*, wherein the presence of ORFX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant ORFX expression or activity.)

The methods of the invention can also be used to detect genetic lesions in a ORFX gene, thereby determining if a subject with the lesioned gene is at risk for, or suffers from, a proliferative disorder, differentiative disorder, glia-associated disorder, etc. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a ORFX-protein, or the mis-expression of the ORFX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of (1) a deletion of one or more nucleotides from a ORFX gene; (2) an addition of one or more nucleotides to a ORFX gene; (3) a substitution of one or more nucleotides of a ORFX gene, (4) a chromosomal

rearrangement of a ORFX gene; (5) an alteration in the level of a messenger RNA transcript of a ORFX gene, (6) aberrant modification of a ORFX gene, such as of the methylation pattern of the genomic DNA, (7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a ORFX gene, (8) a non-wild type level of a ORFX-protein, (9) allelic loss of a ORFX gene, and (10) inappropriate post-translational modification of a ORFX-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a ORFX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *PNAS* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the ORFX-gene (see Abravaya *et al.* (1995) *Nucl Acids Res* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a ORFX gene under conditions such that hybridization and amplification of the ORFX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.*, 1990, *Proc Natl Acad Sci USA* 87:1874-1878), transcriptional amplification system (Kwoh, *et al.*, 1989, *Proc Natl Acad Sci USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, 1988, *BioTechnology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a ORFX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Pat. No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in ORFX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin *et al.* (1996) *Human Mutation* 7: 244-255; Kozal *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in ORFX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al.* above. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the ORFX gene and detect mutations by comparing the sequence of the sample ORFX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert (1977) *PNAS* 74:560 or Sanger (1977) *PNAS* 74:5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve *et al.*, (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publ. No. WO 94/16101; Cohen *et al.* (1996) *Adv Chromatogr* 36:127-162; and Griffin *et al.* (1993) *Appl Biochem Biotechnol* 38:147-159).

Other methods for detecting mutations in the ORFX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA

heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type ORFX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al* (1988) *Proc Natl Acad Sci USA* 85:4397; Saleeba *et al* (1992) *Methods Enzymol* 217:286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in ORFX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a ORFX sequence, *e.g.*, a wild-type ORFX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in ORFX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl Acad Sci USA*: 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control ORFX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting



alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA, rather than DNA, in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. See, e.g., Keen *et al.* (1991) *Trends Genet* 7:5.

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). See, e.g., Myers *et al.* (1985) *Nature* 313:495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, e.g., Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. See, e.g., Saiki *et al.* (1986) *Nature* 324:163; Saiki *et al.* (1989) *Proc Natl Acad. Sci USA* 86:6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. See, e.g., Gasparini *et al.* (1992) *Mol Cell Probes* 6:1. It is anticipated that in certain embodiments amplification may also be performed using Taq ligase

for amplification. See, *e.g.*, Barany (1991) *Proc Natl Acad Sci USA* 88:189. In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

5        The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a ORFX gene.

10       Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which ORFX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

#### **Pharmacogenomics**

15       Agents, or modulators that have a stimulatory or inhibitory effect on ORFX activity (*e.g.*, ORFX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g.*, neurological, cancer-related or gestational disorders) associated with aberrant ORFX activity. In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be  
20       considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate  
25       dosages and therapeutic regimens. Accordingly, the activity of ORFX protein, expression of ORFX nucleic acid, or mutation content of ORFX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

30       Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996, *Clin Exp Pharmacol Physiol*, 23:983-985 and Linder, 1997, *Clin Chem*, 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic

conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of ORFX protein, expression of ORFX nucleic acid, or mutation content of ORFX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a ORFX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

### Monitoring Clinical Efficacy

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied in basic drug screening and in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase ORFX gene expression, protein levels, or upregulate ORFX activity, can be monitored in clinical trials of subjects exhibiting decreased ORFX gene expression, protein levels, or downregulated ORFX activity.

Alternatively, the effectiveness of an agent determined by a screening assay to decrease ORFX gene expression, protein levels, or downregulate ORFX activity, can be monitored in clinical trials of subjects exhibiting increased ORFX gene expression, protein levels, or upregulated ORFX activity. In such clinical trials, the expression or activity of ORFX and, preferably, other genes that have been implicated in, for example, a proliferative or neurological disorder, can be used as a "read out" or marker of the responsiveness of a particular cell.

For example, genes, including ORFX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates ORFX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of ORFX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of ORFX or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, nucleic acid, peptidomimetic, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a ORFX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more

post-administration samples from the subject; (iv) detecting the level of expression or activity of the ORFX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the ORFX protein, mRNA, or genomic DNA in the pre-administration sample with the ORFX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of ORFX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of ORFX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

### Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant ORFX expression or activity.

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, (i) a ORFX polypeptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to a ORFX peptide; (iii) nucleic acids encoding a ORFX peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to a ORFX peptide) that are utilized to "knockout" endogenous function of a ORFX peptide by homologous recombination (see, *e.g.*, Capecchi, 1989, *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between a ORFX peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized

include, but are not limited to, a ORFX peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of a ORFX peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, etc.).

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant ORFX expression or activity, by administering to the subject an agent that modulates ORFX expression or at least one ORFX activity. Subjects at risk for a disease that is caused or contributed to by aberrant ORFX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the ORFX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of ORFX aberrancy, for example, a ORFX agonist or ORFX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

Another aspect of the invention pertains to methods of modulating ORFX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of ORFX protein activity associated with the cell. An agent that modulates ORFX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a ORFX protein, a peptide, a ORFX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more ORFX protein activity. Examples of such stimulatory agents include active ORFX protein and a nucleic acid molecule encoding ORFX that has been introduced into the cell. In another embodiment, the agent inhibits one or more ORFX protein activity. Examples of such inhibitory agents include antisense ORFX nucleic acid molecules and anti-ORFX antibodies. These modulatory methods can be performed *in vitro*

(*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a ORFX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) ORFX expression or activity. In another embodiment, the method involves administering a ORFX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant ORFX expression or activity.

#### **Determination of the Biological Effect of a Therapeutic**

In various embodiments of the present invention, suitable *in vitro* or *in vivo* assays are utilized to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

#### **Malignancies**

Some ORFX polypeptides are expressed in cancerous cells (*see, e.g.*, Tables 1 and 2). Accordingly, the corresponding ORF protein is involved in the regulation of cell proliferation. Accordingly, Therapeutics of the present invention may be useful in the therapeutic or prophylactic treatment of diseases or disorders that are associated with cell hyperproliferation and/or loss of control of cell proliferation (*e.g.*, cancers, malignancies and tumors). For a review of such hyperproliferation disorders, *see e.g.*, Fishman, *et al.*, 1985. MEDICINE, 2nd ed., J.B. Lippincott Co., Philadelphia, PA.

Therapeutics of the present invention may be assayed by any method known within the art for efficacy in treating or preventing malignancies and related disorders. Such assays include,

but are not limited to, *in vitro* assays utilizing transformed cells or cells derived from the patient's tumor, as well as *in vivo* assays using animal models of cancer or malignancies. Potentially effective Therapeutics are those that, for example, inhibit the proliferation of tumor-derived or transformed cells in culture or cause a regression of tumors in animal models, in comparison to the controls.

In the practice of the present invention, once a malignancy or cancer has been shown to be amenable to treatment by modulating (*i.e.*, inhibiting, antagonizing or agonizing) activity, that cancer or malignancy may subsequently be treated or prevented by the administration of a Therapeutic that serves to modulate protein function.

#### **Premalignant conditions**

The Therapeutics of the present invention that are effective in the therapeutic or prophylactic treatment of cancer or malignancies may also be administered for the treatment of pre-malignant conditions and/or to prevent the progression of a pre-malignancy to a neoplastic or malignant state. Such prophylactic or therapeutic use is indicated in conditions known or suspected of preceding progression to neoplasia or cancer, in particular, where non-neoplastic cell growth consisting of hyperplasia, metaplasia or, most particularly, dysplasia has occurred. For a review of such abnormal cell growth see *e.g.*, Robbins & Angell, 1976. BASIC PATHOLOGY, 2nd ed., W.B. Saunders Co., Philadelphia, PA.

Hyperplasia is a form of controlled cell proliferation involving an increase in cell number in a tissue or organ, without significant alteration in its structure or function. For example, it has been demonstrated that endometrial hyperplasia often precedes endometrial cancer. Metaplasia is a form of controlled cell growth in which one type of mature or fully differentiated cell substitutes for another type of mature cell. Metaplasia may occur in epithelial or connective tissue cells. Dysplasia is generally considered a precursor of cancer, and is found mainly in the epithelia. Dysplasia is the most disorderly form of non-neoplastic cell growth, and involves a loss in individual cell uniformity and in the architectural orientation of cells. Dysplasia characteristically occurs where there exists chronic irritation or inflammation, and is often found in the cervix, respiratory passages, oral cavity, and gall bladder.



Alternatively, or in addition to the presence of abnormal cell growth characterized as hyperplasia, metaplasia, or dysplasia, the presence of one or more characteristics of a transformed or malignant phenotype displayed either *in vivo* or *in vitro* within a cell sample derived from a patient, is indicative of the desirability of prophylactic/therapeutic administration of a Therapeutic that possesses the ability to modulate activity of An aforementioned protein. Characteristics of a transformed phenotype include, but are not limited to: (i) morphological changes; (ii) looser substratum attachment; (iii) loss of cell-to-cell contact inhibition; (iv) loss of anchorage dependence; (v) protease release; (vi) increased sugar transport; (vii) decreased serum requirement; (viii) expression of fetal antigens, (ix) disappearance of the 250 kDal cell-surface protein, and the like. See *e.g.*, Richards, *et al.*, 1986. MOLECULAR PATHOLOGY, W.B. Saunders Co., Philadelphia, PA.

In a specific embodiment of the present invention, a patient that exhibits one or more of the following predisposing factors for malignancy is treated by administration of an effective amount of a Therapeutic: (i) a chromosomal translocation associated with a malignancy (*e.g.*, the Philadelphia chromosome (*bcr/abl*) for chronic myelogenous leukemia and t(14;18) for follicular lymphoma, etc.); (ii) familial polyposis or Gardner's syndrome (possible forerunners of colon cancer); (iii) monoclonal gammopathy of undetermined significance (a possible precursor of multiple myeloma) and (iv) a first degree kinship with persons having a cancer or pre-cancerous disease showing a Mendelian (genetic) inheritance pattern (*e.g.*, familial polyposis of the colon, Gardner's syndrome, hereditary exostosis, polyendocrine adenomatosis, Peutz-Jeghers syndrome, neurofibromatosis of Von Recklinghausen, medullary thyroid carcinoma with amyloid production and pheochromocytoma, retinoblastoma, carotid body tumor, cutaneous melanocarcinoma, intraocular melanocarcinoma, xeroderma pigmentosum, ataxia telangiectasia, Chediak-Higashi syndrome, albinism, Fanconi's aplastic anemia and Bloom's syndrome).

In another embodiment, a Therapeutic of the present invention is administered to a human patient to prevent the progression to breast, colon, lung, pancreatic, or uterine cancer, or melanoma or sarcoma.

### **Hyperproliferative and dysproliferative disorders**

In one embodiment of the present invention, a Therapeutic is administered in the therapeutic or prophylactic treatment of hyperproliferative or benign dysproliferative disorders. The efficacy in treating or preventing hyperproliferative diseases or disorders of a Therapeutic of the present invention may be assayed by any method known within the art. Such assays include *in vitro* cell proliferation assays, *in vitro* or *in vivo* assays using animal models of hyperproliferative diseases or disorders, or the like. Potentially effective Therapeutics may, for example, promote cell proliferation in culture or cause growth or cell proliferation in animal models in comparison to controls.

Specific embodiments of the present invention are directed to the treatment or prevention of cirrhosis of the liver (a condition in which scarring has overtaken normal liver regeneration processes); treatment of keloid (hypertrophic scar) formation causing disfiguring of the skin in which the scarring process interferes with normal renewal; psoriasis (a common skin condition characterized by excessive proliferation of the skin and delay in proper cell fate determination); benign tumors; fibrocystic conditions and tissue hypertrophy (e.g., benign prostatic hypertrophy).

### **Neurodegenerative disorders**

Some ORFX proteins are found in cell types have been implicated in the deregulation of cellular maturation and apoptosis, which are both characteristic of neurodegenerative disease. Accordingly, Therapeutics of the invention, particularly but not limited to those that modulate (or supply) activity of an aforementioned protein, may be effective in treating or preventing neurodegenerative disease. Therapeutics of the present invention that modulate the activity of an aforementioned protein involved in neurodegenerative disorders can be assayed by any method known in the art for efficacy in treating or preventing such neurodegenerative diseases and disorders. Such assays include *in vitro* assays for regulated cell maturation or inhibition of apoptosis or *in vivo* assays using animal models of neurodegenerative diseases or disorders, or any of the assays described below. Potentially effective Therapeutics, for example but not by way of limitation, promote regulated cell maturation and prevent cell apoptosis in culture, or reduce neurodegeneration in animal models in comparison to controls.

Once a neurodegenerative disease or disorder has been shown to be amenable to treatment by modulation activity, that neurodegenerative disease or disorder can be treated or prevented by administration of a Therapeutic that modulates activity. Such diseases include all degenerative disorders involved with aging, especially osteoarthritis and neurodegenerative disorders.

#### Disorders related to organ transplantation

Some ORFX can be associated with disorders related to organ transplantation, in particular but not limited to organ rejection. Therapeutics of the invention, particularly those that modulate (or supply) activity, may be effective in treating or preventing diseases or disorders related to organ transplantation. Therapeutics of the invention (particularly Therapeutics that modulate the levels or activity of an aforementioned protein) can be assayed by any method known in the art for efficacy in treating or preventing such diseases and disorders related to organ transplantation. Such assays include *in vitro* assays for using cell culture models as described below, or *in vivo* assays using animal models of diseases and disorders related to organ transplantation, see *e.g.*, below. Potentially effective Therapeutics, for example but not by way of limitation, reduce immune rejection responses in animal models in comparison to controls.

Accordingly, once diseases and disorders related to organ transplantation are shown to be amenable to treatment by modulation of activity, such diseases or disorders can be treated or prevented by administration of a Therapeutic that modulates activity.

#### Cardiovascular Disease

GENX has been implicated in cardiovascular disorders, including in atherosclerotic plaque formation. Diseases such as cardiovascular disease, including cerebral thrombosis or hemorrhage, ischemic heart or renal disease, peripheral vascular disease, or thrombosis of other major vessel, and other diseases, including diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage disease, systemic lupus erythematosus, homocysteinemia, and familial protein or lipid processing diseases, and the like, are either directly or indirectly associated with atherosclerosis. Accordingly, Therapeutics of the invention, particularly those that modulate (or supply) activity or formation may be effective in treating or preventing

atherosclerosis-associated diseases or disorders. Therapeutics of the invention (particularly Therapeutics that modulate the levels or activity) can be assayed by any method known in the art, including those described below, for efficacy in treating or preventing such diseases and disorders.

5           A vast array of animal and cell culture models exist for processes involved in atherosclerosis. A limited and non-exclusive list of animal models includes knockout mice for premature atherosclerosis (Kurabayashi and Yazaki, 1996, *Int. Angiol.* 15: 187-194), transgenic mouse models of atherosclerosis (Kappel *et al.*, 1994, *FASEB J.* 8: 583-592), antisense oligonucleotide treatment of animal models (Callow, 1995, *Curr. Opin. Cardiol.* 10: 569-576),  
10   transgenic rabbit models for atherosclerosis (Taylor, 1997, *Ann. N.Y. Acad. Sci.* 811: 146-152), hypercholesterolemic animal models (Rosenfeld, 1996, *Diabetes Res. Clin. Pract.* 30 Suppl.: 1-11), hyperlipidemic mice (Paigen *et al.*, 1994, *Curr. Opin. Lipidol.* 5: 258-264), and inhibition of lipoxygenase in animals (Sigal *et al.*, 1994, *Ann. N.Y. Acad. Sci.* 714: 211-224). In addition, *in vitro* cell models include but are not limited to monocytes exposed to low density lipoprotein  
15   (Frostegard *et al.*, 1996, *Atherosclerosis* 121: 93-103), cloned vascular smooth muscle cells (Suttles *et al.*, 1995, *Exp. Cell Res.* 218: 331-338), endothelial cell-derived chemoattractant exposed T cells (Katz *et al.*, 1994, *J. Leukoc. Biol.* 55: 567-573), cultured human aortic endothelial cells (Farber *et al.*, 1992, *Am. J. Physiol.* 262: H1088-1085), and foam cell cultures (Libby *et al.*, 1996, *Curr Opin Lipidol* 7: 330-335). Potentially effective Therapeutics, for  
20   example but not by way of limitation, reduce foam cell formation in cell culture models, or reduce atherosclerotic plaque formation in hypercholesterolemic mouse models of atherosclerosis in comparison to controls.

          Accordingly, once an atherosclerosis-associated disease or disorder has been shown to be amenable to treatment by modulation of activity or formation, that disease or disorder can be  
25   treated or prevented by administration of a Therapeutic that modulates activity.

#### **Cytokine and Cell Proliferation/Differentiation Activity**

A GENX protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered

to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D,  
5 DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+ ), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods: Assays for T-cell or thymocyte proliferation include without limitation those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Ed by Coligan *et al.*, Greene Publishing  
10 Associates and Wiley-Interscience (Chapter 3 and Chapter 7); Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Bertagnoli *et al.*, *J Immunol* 145:1706-1712, 1990; Bertagnoli *et al.*, *Cell Immunol* 133:327-341, 1991; Bertagnoli, *et al.*, *J Immunol* 149:3778-3783, 1992; Bowman *et al.*, *J Immunol* 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or  
15 thymocytes include, without limitation, those described by Kruisbeek and Shevach, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1, pp. 3.12.1-14, John Wiley and Sons, Toronto 1994; and by Schreiber, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan eds. Vol 1 pp. 6.8.1-8, John Wiley and Sons, Toronto 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells  
20 include, without limitation, those described by Bottomly *et al.*, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto 1991; deVries *et al.*, *J Exp Med* 173:1205-1211, 1991; Moreau *et al.*, *Nature* 336:690-692, 1988; Greenberger *et al.*, *Proc Natl Acad Sci U.S.A.* 80:2931-2938, 1983; Nordan, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.6.1-5, John Wiley and Sons,  
25 Toronto 1991; Smith *et al.*, *Proc Natl Acad Sci U.S.A.* 83:1857-1861, 1986; Measurement of human Interleukin 11-Bennett, *et al.* In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto 1991; Ciarletta, *et al.*, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto 1991.

30 Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and

cytokine production) include, without limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds., Greene Publishing Associates and Wiley-Interscience (Chapter 3Chapter 6, Chapter 7); Weinberger *et al.*, *Proc Natl Acad Sci USA* 77:6091-6095, 1980; Weinberger *et al.*, *Eur J Immun* 11:405-411, 1981; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988.

### Immune Stimulating or Suppressing Activity

A GENX protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), *e.g.*, in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by vital (*e.g.*, HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by vital, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania species., malaria species. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, *i.e.*, in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response.

The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or energy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon re-exposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7), *e.g.*, preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (*e.g.*, B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this manner prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to energize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the

immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, Science 257:789-792 (1992) and Turka *et al.*, Proc Natl Acad Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., FUNDAMENTAL IMMUNOLOGY, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and auto-antibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of auto-antibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., FUNDAMENTAL IMMUNOLOGY, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the *in vitro* activated T cells into the



patient. Another method of enhancing anti-vital immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells *in vivo*.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (*e.g.*, sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (*e.g.*, a cytoplasmic-domain truncated portion) of an MHC class I  $\alpha$  chain protein and  $\beta_2$  microglobulin protein or an MHC class II  $\alpha$  chain protein and an MHC class II  $\beta$  chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (*e.g.*, B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor

specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods: Suitable assays for thymocyte or splenocyte cytotoxicity include, without  
5 limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 7); Herrmann *et al.*, *Proc Natl Acad Sci USA* 78:2488-2492, 1981; Herrmann *et al.*, *J Immunol* 128:1968-1974, 1982; Handa *et al.*, *J Immunol* 135:1564-1572, 1985; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988; Herrmann *et al.*, *Proc Natl Acad Sci USA* 78:2488-2492, 1981;  
10 Herrmann *et al.*, *J Immunol* 128:1968-1974, 1982; Handa *et al.*, *J Immunol* 135:1564-1572, 1985; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Bowman *et al.*, *J Virology* 61:1992-1998; Takai *et al.*, *J Immunol* 140:508-512, 1988; Bertagnolli *et al.*, *Cell Immunol* 133:327-341, 1991; Brown *et al.*, *J Immunol* 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which  
15 will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, *J Immunol* 144:3028-3033, 1990; and Mond and Brunswick In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, (eds.) Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins  
20 that generate predominantly Th1 and CTL responses) include, without limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 7); Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988; Bertagnolli *et al.*, *J Immunol* 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by  
25 dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery *et al.*, *J Immunol* 134:536-544, 1995; Inaba *et al.*, *J Exp Med* 173:549-559, 1991; Macatonia *et al.*, *J Immunol* 154:5071-5079, 1995; Porgador *et al.*, *J Exp Med* 182:255-260, 1995; Nair *et al.*, *J Virol* 67:4062-4069, 1993; Huang *et al.*, *Science* 264:961-965, 1994; Macatonia *et al.*, *J Exp Med* 169:1255-1264, 1989; Bhardwaj *et al.*, *J Clin Investig* 94:797-807, 1994; and Inaba *et al.*, *J Exp Med* 172:631-640, 1990.  
30

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz *et al.*, *Cytometry* 13:795-808, 1992; Gorczyca *et al.*, *Leukemia* 7:659-670, 1993; Gorczyca *et al.*, *Cancer Res* 53:1945-1951, 1993; Itoh *et al.*, *Cell* 66:233-243, 1991; Zacharchuk, *J Immunol* 145:4037-4045, 1990; Zamai *et al.*, *Cytometry* 14:891-897, 1993; Gorczyca *et al.*, *Internat J Oncol* 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica *et al.*, *Blood* 84:111-117, 1994; Fine *et al.*, *Cell Immunol* 155: 111-122, 1994; Galy *et al.*, *Blood* 85:2770-2778, 1995; Toki *et al.*, *Proc Nat Acad Sci USA* 88:7548-7551, 1991.

#### **Hematopoiesis Regulating Activity**

A GENX protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, *e.g.* in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (*i.e.*, traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (*i.e.*, in conjunction with bone marrow

transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

5           Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson *et al.* *Cellular Biology* 15:141-151, 1995; Keller *et al.*, *Mol. Cell.*  
10 *Biol.* 13:473-486, 1993; McClanahan *et al.*, *Blood* 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* (eds.) Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama *et al.*,  
15 *Proc Natl Acad Sci USA* 89:5907-5911, 1992; McNiece and Briddeli, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* (eds.) Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben *et al.*, *Exp Hematol* 22:353-359, 1994; Ploemacher, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Spoonceret *et al.*, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.*, (eds.) Vol pp.  
20 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Sutherland, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.*, (eds.) Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

#### **Tissue Growth Activity**

A GENX protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for  
25 wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation  
30 employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation

induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendonitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, *i.e.* for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein

may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, EPIDERMAL WOUND HEALING, pp. 71-112 (Maibach and Rovee, eds.), Year Book Medical

Publishers, Inc., Chicago, as modified by Eaglstein and Menz, *J. Invest. Dermatol* 71:382-84 (1978).

#### **Activin/Inhibin Activity**

5 A GENX protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in  
10 male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-b group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A protein of the invention may  
15 also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

20 Assays for activin/inhibin activity include, without limitation, those described in: Vale *et al.*, *Endocrinology* 91:562-572, 1972; Ling *et al.*, *Nature* 321:779-782, 1986; Vale *et al.*, *Nature* 321:776-779, 1986; Mason *et al.*, *Nature* 318:659-663, 1985; Forage *et al.*, *Proc Natl Acad Sci USA* 83:3091-3095, 1986.

#### **Chemotactic/Chemokinetic Activity**

25 A protein of the present invention may have chemotactic or chemokinetic activity (*e.g.*, act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of  
30 wounds and other trauma to tissues, as well as in treatment of localized infections. For example,

attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Coligan *et al.*, eds. (Chapter 6.12, MEASUREMENT OF ALPHA AND BETA CHEMOKINES 6.12.1-6.12.28); Taub *et al. J Clin Invest* 95:1370-1376, 1995; Lind *et al. APMIS* 103:140-146, 1995; Muller *et al., Eur J Immunol* 25: 1744-1748; Gruber *et al. J Immunol* 152:5860-5867, 1994; Johnston *et al., J Immunol* 153: 1762-1768, 1994.

#### **Hemostatic and Thrombolytic Activity**

A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (*e.g.*, stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet *et al., J. Clin. Pharmacol.* 26:131-140, 1986; Burdick *et al., Thrombosis Res.*



45:413-419, 1987; Humphrey *et al.*, *Fibrinolysis* 5:71-79 (1991); Schaub, *Prostaglandins* 35:467-474, 1988.

### **Receptor/Ligand Activity**

A protein of the present invention may also demonstrate activity as receptors, receptor  
5 ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and  
ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and  
their ligands, receptor phosphatases and their ligands, receptors involved in cell—cell  
interactions and their ligands (including without limitation, cellular adhesion molecules (such as  
10 selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation,  
antigen recognition and development of cellular and humoral immune responses). Receptors and  
ligands are also useful for screening of potential peptide or small molecule inhibitors of the  
relevant receptor/ligand interaction. A protein of the present invention (including, without  
limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of  
receptor/ligand interactions.

15 The activity of a protein of the invention may, among other means, be measured by the  
following methods:

Suitable assays for receptor-ligand activity include without limitation those described in:  
CURRENT PROTOCOLS IN IMMUNOLOGY, Ed by Coligan, *et al.*, Greene Publishing Associates and  
Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions  
20 7.28.1-7.28.22), Takai *et al.*, *Proc Natl Acad Sci USA* 84:6864-6868, 1987; Bierer *et al.*, *J. Exp.*  
*Med.* 168:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169:149-160 1989; Stoltenborg *et al.*,  
*J Immunol Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670, 1995.

### **Anti-Inflammatory Activity**

Proteins of the present invention may also exhibit anti-inflammatory activity. The  
25 anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the  
inflammatory response, by inhibiting or promoting cell—cell interactions (such as, for example,  
cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory  
process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production  
of other factors which more directly inhibit or promote an inflammatory response. Proteins  
30 exhibiting such activities can be used to treat inflammatory conditions including chronic or acute

conditions), including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

#### **Tumor Inhibition Activity**

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

#### **Other Activities**

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting

deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

Neural disorders in general include Parkinson's disease, Alzheimer's disease, Huntington's disease, multiple sclerosis, amyotrophic lateral sclerosis (ALS), peripheral neuropathy, tumors of the nervous system, exposure to neurotoxins, acute brain injury, peripheral nerve trauma or injury, and other neuropathies, epilepsy, and/or tremors.

## **EQUIVALENTS**

From the foregoing detailed description of the specific embodiments of the invention, it should be apparent that particular novel compositions and methods involving nucleic acids, polypeptides, antibodies, detection and treatment have been described. Although these particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims that follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made as a matter of routine for a person of ordinary skill in the art to the invention without departing from the spirit and scope of the invention as defined by the claims. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

**THIS PAGE INTENTIONALLY LEFT BLANK**

**THIS PAGE INTENTIONALLY LEFT BLANK**

**THIS PAGE INTENTIONALLY LEFT BLANK**

**THIS PAGE INTENTIONALLY LEFT BLANK**

**THIS PAGE INTENTIONALLY LEFT BLANK**



**THIS PAGE INTENTIONALLY LEFT BLANK**

**THIS PAGE INTENTIONALLY LEFT BLANK**

**THIS PAGE INTENTIONALLY LEFT BLANK**

Table 1

ORF#	Internal Identification Number	Protein similarity	Protein domain	Protein Classification	Cells or Tissues in which Gene is Expressed
1	13076366 (1, 2)	Novel Protein sim. GBank gi 4691395 emb CAB41562.1  - (AL049727) putative large secreted protein [Streptomyces coelicolor]		UNCLASSIFIED	264636
2	80248091 (3, 4)	Novel Protein sim. GBank gi 2829506 sp P71559 SUCC_MYCTU - SUCCINYL-COA SYNTHETASE BETA CHAIN (SCS-BETA)	Contains protein domain (PF00549) - CoA-ligases	UNCLASSIFIED	264807, 264600, 264602, 264762, 264769, 264689, 264638, 264567
3	80415924 (5, 6)			UNCLASSIFIED	264910, 264604, 264634, 264905, 264636, 264691, 264907, 264692, 264629
4	82018837 (7, 8)			UNCLASSIFIED	264908, 264909, 264760, 264628, 264635
5	79970035 (9, 10)			UNCLASSIFIED	22279002, 264563
6	79842462 (11, 12)		Contains protein domain (PF00127) - Copper binding proteins, plastocyanin/azurin family	UNCLASSIFIED	264908
7	85515576 (13, 14)	Novel Protein sim. GBank gi 415928 gb AAD20157  - (AC006282) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	20281099, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 265006, 264512, 265009, 264910, 264595, 264596, 264758, 264603, 264604, 264760, 264762, 264683, 264768, 264767, 264689, 35695917, 264890, 264892, 264693, 33657109, 264628, 264629, 35696423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264636, 264637, 264638, 264639, 18108385, 264563, 264564, 264566, 264486
8	56924276 (15, 16)	Novel Protein sim. GBank gi 585562 sp Q06458 NIRB_KLEPN - NITRITE REDUCTASE (NAD(P)H) LARGE SUBUNIT		reductase	264907
9	79394457 (17, 18)			UNCLASSIFIED	265007, 265019, 263972
10	78559459 (19, 20)			UNCLASSIFIED	264908
11	20414027 (21, 22)				264605
12	94141210 (23, 24)	Novel Protein sim. GBank gi 3878145 emb CAA99871  - (Z75543) similar to potassium channel protein [Caenorhabditis elegans]		misc_channel	264259, 265007, 83373044
13	20750551 (25, 26)			UNCLASSIFIED	264556, 264557, 264564
14	95105114 (27, 28)	Novel Protein sim. GBank gi 283278 emb CAA12845  - (AJ225805) inward potassium channel alpha subunit [Egeria densa]	Contains protein domain (PF00023) - Atk repeat	potassium_channel	35696286, 35696052, 264510, 35695917, 264691, 264628, 35696423, 264555, 264558, 264559, 83373044
15	20458307 (29, 30)	Novel Protein sim. GBank gi 1710791 sp Q10234 RT05_SCHPO - PROBABLE MITOCHONDRIAL 40S RIBOSOMAL PROTEIN S5	Contains protein domain (PF00333) - Ribosomal protein S5	ribosomalprot	264604
16	20760356 (31, 32)				264555

17	20292744 (33, 34)	Novel Protein sim. GBank gij174884[spP4439]URE1_HAEIN - UREASE ALPHA SUBUNIT (UREA AMIDOHYDROLASE)	Contains protein domain (PF00449) - Urease		264600
18	80246804 (35, 36)	Novel Protein sim. GBank gij2281102 (AC002333) - SF16 isolog [Arabidopsis thaliana]			29331827, 264555, 264557, 264638, 264558
19	80076624 (37, 38)	Novel Protein sim. GBank gij2506112[spP43672]UUP_ECOLI - ABC TRANSPORTER ATP-BINDING PROTEIN UUP	UNCLASSIFIED transport		22278996, 264907, 264910, 264600, 264693 264602
20	20724558 (39, 40)	Novel Protein sim. GBank gij1730203[spP50442]GATM_RAT - GLYCINE AMIDINOTRANSFERASE PRECURSOR (L- ARGININE:GLYCINE AMIDINOTRANSFERASE) (TRANSAMIDINASE) (AT)	UNCLASSIFIED		22278995, 264906, 265008, 265010, 265011, 264602, 264605, 264768, 264688, 21806784, 264691, 18108376, 264638, 18108387, 264486
21	80417554 (41, 42)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase		264685 264488, 264907, 264909, 264600, 264602, 264603, 264605, 264682, 264768, 32833986, 264636, 264486
22	11705858 (43, 44)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]			264600
23	80418176 (45, 46)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]			264593
24	20291697 (47, 48)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]			22278996, 56182435, 265018, 264566
25	80253774 (49, 50)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	UNCLASSIFIED ribosomal prot		18108370, 35696423, 264635, 264555
26	80255394 (51, 52)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]			264638
27	80235795 (53, 54)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	Contains protein domain (PF00253) - Ribosomal protein S14p/S29e		264601, 264762, 264766, 264769, 264638
28	79483561 (55, 56)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	UNCLASSIFIED kinase		264638
29	82448765 (57, 58)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	Contains protein domain (PF00365) - Phosphofructokinase		264601, 264762, 264766, 264769, 264638
30	79169333 (59, 60)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	UNCLASSIFIED		264908, 265019, 264687, 21906764, 21906766
31	19848158 (61, 62)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	UNCLASSIFIED		264534
32	82449485 (63, 64)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	UNCLASSIFIED		264905, 264605, 264762, 264766, 264687, 264689
33	79582628 (65, 66)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	UNCLASSIFIED		264687
34	87467657 (67, 68)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	UNCLASSIFIED		60432289, 264600, 264602, 264760, 18108357, 264769, 265020, 264691
35	95005170 (69, 70)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	UNCLASSIFIED		264600, 264687, 264558, 264639
36	19842042 (71, 72)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	synthase		264566
37	20369215 (73, 74)	Novel Protein sim. GBank gij187329[emb]CAB07071 - (Z82771) fadE25 [Mycobacterium tuberculosis]	dehydrogenase		264603

38	20466334 (75, 76)	Novel Protein sim. GBank gij3805970[emb]CAA06231] - (AJ004933) periplasmic nitrate reductase, large subunit [Rhodospseudomonas sp.]		reductase	264605
39	94300715 (77, 78)	Novel Protein sim. GBank gij192449 (L63543) - endodermin [Xenopus laevis]	Contains protein domain (PF00207) - Alpha-2-macroglobulin family	complement	264805, 264806, 264807, 66712502, 264808, 264909, 264511, 265009, 264910, 55812036, 264758, 265011, 264762, 264682, 264783, 264764, 264766, 265022, 264693, 264628, 264631, 264634, 264635, 264555, 264638, 18108381, 264558, 18108385, 264482
40	20635625 (79, 80)			UNCLASSIFIED	264592
41	80023287 (81, 82)	Novel Protein sim. GBank gij854065[emb]CAA56337] - (X83413) U88 [Human herpesvirus 6]			264591, 35695917
42	20724566 (83, 84)			UNCLASSIFIED	264602
43	20467089 (85, 86)	Novel Protein sim. GBank gij3820584 (AF086781) - carbamoylphosphate synthetase large subunit [Zymomonas mobilis]		synthase	264605
44	13085287 (87, 88)	Novel Protein sim. GBank gij2494784[sp]Q50729[GUA, MYCTU - GMP SYNTHASE (GLUTAMINE-HYDROLYZING) (GLUTAMINE AMIDOTRANSFERASE) (GMP SYNTHETASE)	Contains protein domain (PF00958) - GMP synthase C terminal domain	synthase	264769, 264536
45	36384711 (89, 90)	Novel Protein sim. GBank gij1881738 (U89688) - myosin-I binding protein Acan125 [Acanthamoeba castellanii]		UNCLASSIFIED	264769, 264510, 264508
46	95003388 (91, 92)			ngl	264566
47	11698624 (93, 94)			UNCLASSIFIED	264889
48	79407218 (95, 96)				18108385, 264635, 264828
49	21659844 (97, 98)			UNCLASSIFIED	264603
50	80503996 (99, 100)			UNCLASSIFIED	264508, 264603, 264769, 264689, 264636, 264558, 264486
51	80235569 (101, 102)	Novel Protein sim. GBank gij3411177 (AF076240) - MocC [Rhizobium leguminosarum bv. viciae]		UNCLASSIFIED	264593, 18108387
52	78208528 (103, 104)	Novel Protein sim. GBank gij3914992[sp]Q26264[SM41_HEMPU - 41 KD SPICULE MATRIX PROTEIN PRECURSOR (HSM41) (HPSMC)]		sinud	264634
53	36966970 (105, 106)	Novel Protein sim. GBank gij3880411 (AC004561) - putative proline-rich protein [Arabidopsis thaliana]		UNCLASSIFIED	264782
54	79570897 (107, 108)			UNCLASSIFIED	264630, 264909, 264766
55	80202703 (109, 110)	Novel Protein sim. GBank gij1633572 (J52064) - Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus]		UNCLASSIFIED	29331824, 264102, 265018, 18108378
56	8758408 (111, 112)	Novel Protein sim. GBank gij4321580[sp]AAD15785] - (AF050114) alginate lyase [Pseudomonas sp. W7]			264604
57	11223386 (113, 114)		Contains protein domain (PF00076) - RNA recognition motif, (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264557

58	91227506 (115, 116)	Novel Protein sim. GBank gi15616074 gb AAD5616.1 AF06194 - (AF06194) protate- derived STE20-like kinase PSK [Homo sapiens]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	56182575, 264259, 60432049, 35596052, 66712502, 264909, 265008, 265010, 265011, 264881, 29148784, 35695917, 60170815, 264891, 264692, 264693, 18108374, 35698423, 56182323, 60432113 264600, 264689, 264638
59	80077371 (117, 118)	Novel Protein sim. GBank gi1172920 sp P45830 PFE_MYCLE - PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N- ACETYLGLUCOSAMINYLTTRANSFERASE	Contains protein domain (PF00943) - Glycosyl transferase	transferase	
60	12955341 (119, 120)	Novel Protein sim. GBank gi11710215 (U79260) - unknown [Homo sapiens]			264689
61	80426806 (121, 122)			glycoprotein	264766
62	13504866 (123, 124)				264630
63	16474553 (125, 126)			UNCLASSIFIED	265019
64	20724578 (127, 128)	Novel Protein sim. GBank gi122312 sp O06134 KPYK_MYCTU - PYRUVATE transposase homolog (insertion element ISAE1) - Alcaligenes eutrophus		UNCLASSIFIED	264602
65	79326305 (129, 130)	Novel Protein sim. GBank gi13122312 sp O06134 KPYK_MYCTU - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate kinase	kinase	264563
66	46854384 (131, 132)	Novel Protein sim. GBank gi13122312 sp O06134 KPYK_MYCTU - PYRUVATE KINASE (PK)		transport	22278996, 264558
67	78952543 (133, 134)	Novel Protein sim. GBank gi131985 sp P30234 DHA_MYCTU - ALANINE DEHYDROGENASE (40 KD ANTIGEN)		dehydrogenase	265021
68	79817382 (135, 136)				264909
69	79841764 (137, 138)			UNCLASSIFIED	264908
70	79871329 (139, 140)				264906, 264908
71	65897456 (141, 142)			UNCLASSIFIED	264602, 265021
72	87734977 (143, 144)	Novel Protein sim. GBank gi4415928 gb AAD201571 - (AC006282) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	264488, 264905, 264906, 264907, 264908, 264511, 265008, 264910, 264758, 87168474, 264882, 264766, 264886, 264889, 35695917, 265021, 60170815, 264891, 33657023, 264692, 264693, 264629, 264631, 264639, 22279000
73	80025241 (145, 146)			UNCLASSIFIED	60424178, 264508, 264908, 265007, 264603, 264687, 264689, 264692, 18108387
74	20377410 (147, 148)				264605
75	11819032 (149, 150)	Novel Protein sim. GBank gi12853098 emb CAA16914 - (AL021767) vacuolar protein sorting [Schizosaccharomyces pombe]		UNCLASSIFIED	264689
76	95105303 (151, 152)	Novel Protein sim. GBank gi14468811 emb CAB38212 - (AL035601) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	83373044, 264906, 264557
77	10144718 (153, 154)	Novel Protein sim. GBank gi1834065 emb CAA583371 - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264563
78	8758258 (155, 156)			UNCLASSIFIED	264604

79	84140180 (157, 158)	Novel Protein sim. GBank gij5689453jdbjBAA83010.1) - (AB028981) KIAA1058 protein [Homo sapiens]	Contains protein domain (PF00169) - PH domain	35696286, 22278988, 29331822, 29331824, 29331825, 29331827, 264905, 264906, 264907, 66712502, 264908, 264909, 265008, 265009, 264910, 60170831, 55812038, 33109954, 265017, 265018, 264288, 264768, 56181562, 21906765, 21906769, 29148784, 265020, 264690, 264691, 264692, 264693, 60431528, 35696423, 264631, 264632, 264634, 264638, 264639, 83373044, 264564, 264566, 264567
80	82314840 (159, 160)		UNCLASSIFIED	264769, 264601, 265006, 264910, 264604, 264605, 264634, 264635, 264905, 264782, 264637, 264592, 264628, 264907, 264691, 264908, 264567, 264609, 264766
81	20467247 (161, 162)	Novel Protein sim. GBank gij1723442lsp10258jYD2A_SCHPO - HYPOTHETICAL 69.0 KD PROTEIN C56F8.10 IN CHROMOSOME 1	reductase	264605
82	16331388 (163, 164)	Novel Protein sim. GBank gij2855866 (AF045770) - methylmalonate semi-aldohyde dehydrogenase [Oryza sativa]	dehydrogenase	264557
83	94741180 (165, 166)	Novel Protein sim. GBank gij3402673 (AC004697) - unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	264488, 264508, 264509, 264905, 264908, 264909, 264511, 264591, 264593, 264594, 264595, 264598, 264758, 264603, 264760, 264681, 18108351, 264762, 264682, 264764, 264684, 264766, 264686, 264632, 264637, 264557, 264638, 264639, 18108385, 264568
84	80355375 (167, 168)	Novel Protein sim. GBank gij173364lsp145080jSAT1_RAT - SULFATE ANION TRANSPORTER 1 (CANALICULAR SULFATE TRANSPORTER) (SULFATE/CARBONATE ANTI-PORTER)	transport	264508, 264908, 264907, 264909, 264909, 264910, 264760, 264763, 264764, 264786, 264788, 264789, 35695855, 264638, 264637
85	80496600 (169, 170)	Novel Protein sim. GBank gij2120998lprj1S70682 - glycosyltransferase homolog - Bordetella pertussis	transferase	264605, 264762, 264687, 264769, 18108374, 264636, 264486
86	39559043 (171, 172)	Novel Protein sim. GBank gij3256023lmbjCAA17228.1) - (AL021897) hypothetical protein Rv1112 [Mycobacterium tuberculosis]		264910
87	13656808 (173, 174)		UNCLASSIFIED	264093



88	95344718 (175, 176)	Novel Protein sim. GBank gij559703[dbj BAA07552] - (D38549) ha1025 is new [Homo sapiens]	52644507, 52646365, 18108398, 65274572, 56182575, 56984075, 35696286, 22278997, 22278998, 22278999, 264082, 264093, 264094, 264095, 264259, 29331822, 29331824, 56182181, 56714117, 29331825, 29331826, 60432289, 29331827, 29331828, 35696052, 33656970, 264105, 264508, 264905, 264906, 264907, 264908, 29331830, 56712502, 52644045, 56182435, 265007, 265009, 264910, 60170831, 264592, 60431735, 60433356, 33657402, 264757, 60433438, 55812038, 264758, 21906754, 52646317, 33109954, 52644298, 87188474, 265011, 87188559, 264601, 285017, 285018, 264604, 265019, 264448, 264369, 264288, 264766, 52644229, 21906768, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 60170615, 52644150, 33657023, 65274620, 33657109, 27486261, 27486264, 33657349, 35695763, 264628, 263972, 18108374, 55810764, 35696423, 55811576, 65274791, 35695855, 60431650, 264636, 52644332, 56182323, 60170394, 83373044, 18108385, 18108387, 18108388, 56526486, 87188518, 60432113, 22279000, 22279002, 264482, 264564, 264486
89	80077389 (177, 178)	Novel Protein sim. GBank gij1710383[sp P46352 RIPX_BACSU - PROBABLE INTEGRASE/RECOMBINASE RIPX	264600
90	82115999 (179, 180)	Novel Protein sim. GBank gij2498891[sp P76403 YEGQ_ECOLI - PUTATIVE PROTEASE IN BAER-OGRK INTERGENIC REGION	UNCLASSIFIED protease
91	78906950 (181, 182)	Novel Protein sim. GBank gij3367754[emb CAA20079] - (AL031155) hypothetical protein SC3A7.16c [Streptomyces coelicolor]	264760 265006
92	79554871 (183, 184)	Novel Protein sim. GBank gij2895095 (AF011337) - putative E1-E2 ATPase [Mus musculus]	UNCLASSIFIED 264691
93	80495778 (185, 186)	Novel Protein sim. GBank gij1171919[sp P46920 OPUA_BACSU - GLYCINE BETAINES TRANSPORT ATP-BINDING PROTEIN OPUA	264907, 264908, 264910, 265009, 264605, 264769 264906
94	79646649 (187, 188)	Novel Protein sim. GBank gij1171919[sp P46920 OPUA_BACSU - GLYCINE BETAINES TRANSPORT ATP-BINDING PROTEIN OPUA	Contains protein domain (PF00571) - transport CBS domain
95	11090238 (189, 190)		264594

96	94322125 (191, 192)	Novel Protein sim. GBank gii4589560[dj]BAA76802.1] - (AB023175) KIAA0958 protein [Homo sapiens]	UNCLASSIFIED	22278995, 22278989, 264259, 29331822, 29331826, 35696052, 29146499, 264509, 264906, 264907, 264908, 264909, 265007, 265008, 264910, 265009, 264593, 265010, 265017, 264604, 265019, 18108351, 264288, 264786, 264788, 264769, 21906765, 21906767, 21906769, 265020, 264692, 33657182, 35695763, 264628, 264629, 18108379, 264631, 264636, 18108381, 264559, 18108382, 83373044, 22278902
97	79605200 (193, 194)	Novel Protein sim. GBank gii4583559[emb]CAB40388.1] - (AJ005255) OxyR [Erwinia chrysanthemi]	UNCLASSIFIED	264508
98	79427000 (195, 196)	Novel Protein sim. GBank gii1007693[dj]BAA10430] - (D64002) hypothetical protein [Synchocystis sp.]	UNCLASSIFIED	264909
99	20466524 (197, 198)	Novel Protein sim. GBank gii169479[ep]P43825[EF]G_HAEN - ELONGATION FACTOR G (EF-G)	UNCLASSIFIED	264805
100	79640113 (199, 200)	Novel Protein sim. GBank gii480897[pri]S37485 - gene msg1 protein - mouse	UNCLASSIFIED	264693
101	80203208 (201, 202)	Novel Protein sim. GBank gii2894166[emb]CAA11773.1] - (AJ223998) PCZA361.18 [Amycolatopsis orientalis]	UNCLASSIFIED	265020, 264102, 263972
102	20467259 (203, 204)	Novel Protein sim. GBank gii1731040[sp]P54509[QHH_BACSU - HYPOTHETICAL HELICASE IN SINI-GCVT INTERGENIC REGION	synthase	264605
103	20465388 (205, 206)	Novel Protein sim. GBank gii4887229[gb]AAD3244.1AF15075 - (AF150755) microtubule-actin crosslinking factor [Mus musculus]	helicase	264605
104	80247572 (207, 208)	Novel Protein sim. GBank gii854065[emb]CAA58337] - (X83413) U88 [Human herpesvirus 8]	UNCLASSIFIED	264591, 264595, 264602
105	79605206 (209, 210)	Novel Protein sim. GBank gii1685117 (U70770) - furrowed [Drosophila melanogaster]	complement	264508
106	28382058 (211, 212)	Novel Protein sim. GBank gii1705505[sp]P34729[BS4_MOUSE - BS4 PROTEIN	UNCLASSIFIED	264511, 265009
107	80057791 (213, 214)	Novel Protein sim. GBank gii4887229[gb]AAD3244.1AF15075 - (AF150755) microtubule-actin crosslinking factor [Mus musculus]	ATPase-associated	29331824, 264591, 21906754, 265019
108	80237936 (215, 216)	Novel Protein sim. GBank gii283577[emb]CAB15264] - (Z99120) similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]	transport	18108374, 35695917, 22278996, 264113, 264600, 264602, 264603, 265017, 264910, 264908, 264636, 264786
109	95194148 (217, 218)	Novel Protein sim. GBank gii233079[emb]CAB11265] - (Z98801) carboxypeptidase s precursor [Schizosaccharomyces pombe]	UNCLASSIFIED	264758, 264803, 264630, 264636, 264637
110	79582823 (219, 220)			264687
111	39565458 (221, 222)			264564
112	78650338 (223, 224)		UNCLASSIFIED	264908
113	17959439 (225, 226)		UNCLASSIFIED	265007
114	80502101 (227, 228)		UNCLASSIFIED	264769

115	80251003 (229, 230)	Novel Protein sim. GBank gi 2246532 (U93872) - ORF 73; contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		UNCLASSIFIED	52845156, 52845080, 33656970, 284592, 21908754, 27486264, 18108379, 35696423, 284635, 52644332, 18108382, 284905, 284906, 284907, 284908, 284910, 284758, 285010, 284763, 284682, 284764, 284768, 284885, 284686, 284768, 284789, 33657023, 284693, 33657108, 284628, 18108374, 284631, 284632, 284634, 284636, 284637, 284638, 284639, 56526486, 284565, 284566, 284639, 284653
116	81288689 (231, 232)				
117	79636695 (233, 234)		Contains protein domain (PF00170) - bZIP transcription factor	UNCLASSIFIED	263974
118	80222170 (235, 236)				
119	91013071 (237, 238)	Novel Protein sim. GBank gi 732526 (U22327) - alpha2(IV) collagen [Caenorhabditis elegans]		UNCLASSIFIED	22278996, 28331824, 60432289, 265007, 60433438, 264603, 264605, 18108351, 264769, 264889, 265020, 264534, 27486261, 264558, 83373044, 18108385, 284964
120	8756491 (239, 240)	Novel Protein sim. GBank gi 2131219ipr S50157 - cyclin-dependent kinase chain SRB10 - yeast (Saccharomyces cerevisiae)		kinase	264603
121	80026153 (241, 242)				264595
122	20457620 (243, 244)	Novel Protein sim. GBank gi 2052147ipr CAB08137 - (Z84752) kscA [Mycobacterium tuberculosis]	Contains protein domain (PF00398) - Ribosomal RNA adenine dimethylases	transferase	284605
123	8758278 (245, 246)	Novel Protein sim. GBank gi 2833385ipr Q43134 UGST_SORBI - GRANULE-BOUND GLYCAGEN (STARCH) SYNTHASE PRECURSOR			284604
124	79104017 (247, 248)	Novel Protein sim. GBank gi 475542 (U08255) - glutamate receptor delta-1 subunit [Rattus norvegicus]		synthase	18108394, 18108397, 265005, 265007, 265008, 265010, 265011, 18108355, 18108378, 18108380, 18108384
125	87797988 (249, 250)	Novel Protein sim. GBank gi 5102785ipr CAB45200.1 - (AL079308) putative transcriptional regulator [Streptomyces coelicolor]	Contains protein domain (PF00060) - Ligand-gated ion channel	misc_channel	284908, 264906, 265009, 264596, 22278002
126	56701283 (251, 252)				264511
127	20467267 (253, 254)	Novel Protein sim. GBank gi 130120ipr P23620 PHOB_PSEAE - PHOSPHATE REGULON TRANSCRIPTIONAL REGULATORY PROTEIN PHOB	Contains protein domain (PF00072) - Response regulator receiver domain	UNCLASSIFIED - phosphatase	264605
128	80248473 (255, 256)				284907, 284909, 284910, 264600, 264601, 284603, 284605, 18108351, 284693, 284557
129	95290543 (257, 258)	Novel Protein sim. GBank gi 2506493ipr P38036 YCCB_ECOLI - HYPOTHETICAL 100.5 KD PROTEIN IN IAP-CYSH INTERGENIC REGION	Contains protein domain (PF00270) - DEAD/DEAH box helicase	UNCLASSIFIED	35696423, 35695855, 284600, 284602, 284603, 264604, 264605, 284508, 284908, 284584, 264628, 284682, 284565, 284683
130	80085583 (259, 260)	Novel Protein sim. GBank gi 854065ipr CAA58371 - (X83413) U88 [Human herpesvirus 6]			284634
131	94995022 (261, 262)	Novel Protein sim. GBank gi 1076038ipr S54860 - ABC transporter PstC-2 chain - Mycobacterium tuberculosis	Contains protein domain (PF00528) - Binding-protein-dependent transport systems inner membrane component	transport	18108376, 284769, 29331826, 264689, 22278996, 265021, 284600, 264511, 264601, 284602, 264605, 284905, 284636

132	10887692 (263, 264)	Novel Protein sim. GBank gll1877340[emb]CAB07068] - (Z92771) accA3 [Mycobacterium tuberculosis]	Contains protein domain (PF00269) - Carbamoyl-phosphate synthase (CPSase)	carboxylase	264636 264905, 264689
133	84630883 (265, 266)	Novel Protein sim. GBank gll4585838[emb]CAB40932.1] - (AL049630) putative NADH dehydrogenase [Streptomyces coelicolor]		dehydrogenase	264905, 264605, 265021
134	79834880 (267, 268)	Novel Protein sim. GBank gll1460074[emb]CAB01049] - (Z77260) hypothetical protein Rv2566 [Mycobacterium tuberculosis]			264634
135	19885057 (269, 270)	Novel Protein sim. GBank gll2125896[emb]CAA73511] - (Y13070) foxy/polyglutamate synthase [Streptomyces coelicolor]		synthase	264308
136	79846083 (271, 272)	Novel Protein sim. GBank gll5889912[emb]CAB52075.1] - (AL109732) putative mutase [Streptomyces coelicolor A3(2)]	Contains protein domain (PF01817) - Chorismate mutase	dehydrogenase	264683, 264685, 264686, 264691, 264692, 264693 18108374, 18108385, 33857109, 33857182, 265010, 22278998, 265008, 265007, 265008, 265009, 264683
137	79819770 (273, 274)	Novel Protein sim. GBank gll3581916[emb]CAA20855] - (AL031545) mutS family DNA mismatch repair protein [Schizosaccharomyces pombe]		UNCLASSIFIED	264908 264600 264602, 265017
138	79835971 (275, 276)	Novel Protein sim. GBank gll3334127[sp]P97303BAC2_MOUSE - TRANSCRIPTION REGULATOR PROTEIN BACH2 (BTB AND CNC HOMOLOG 2)		UNCLASSIFIED	264600 263978, 264600, 264910, 264632, 264508, 264563, 264564, 264591, 264556, 264809, 264629, 264639
139	86688076 (277, 278)	Novel Protein sim. GBank gll3757569[emb]CAA21315] - (AL031863) 1-evidence=predicted by content; 1-method=genefinder.084; 1-method_score=66.31; 1-evidence_end [Drosophila melanogaster]		UNCLASSIFIED	22278998, 264602
140	79825759 (278, 280)	Novel Protein sim. GBank gll1408071[sp]P24536[Y121_BURCE - INSERTION ELEMENT IS402 HYPOTHETICAL 24 KD PROTEIN]	Contains protein domain (PF01675) - Transposase		264687
141	20700094 (281, 282)	Novel Protein sim. GBank gll2827608[emb]CAA16663] - (AL021648) uvrD2 [Mycobacterium tuberculosis]		helicase	264905, 264906, 264809, 264310, 265009, 60433358, 264600, 264601, 264604, 264605, 264687, 264769, 18108365, 65274791, 18108387
142	80028104 (283, 284)	Novel Protein sim. GBank gll2916947[emb]CAA17585] - (AL021899) hypothetical protein Rv0986 [Mycobacterium tuberculosis]		transport	265009, 265010, 264600, 264602, 264603, 264604, 264605, 264693, 33657109, 264636

149	80249373 (287, 288)	Novel Protein sim. GBank gii1723073 sp Q11040 Y081_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.01	Contains protein domain (PF00005) - transport ABC transporter	265010, 264600, 264601, 264603, 264604, 27486265, 264636
150	20294748 (298, 300)	Novel Protein sim. GBank gii3724125 emb CAA11905  - (AJ224340) maltosephosphorylase [Lactobacillus sanfrancisco]		264600
151	20726388 (301, 302)	Novel Protein sim. GBank gii729312 sp P07651 DEOB_ECOLI - PHOSPHOPENTOMUTASE (PHOSPHODEOXYRIBOMUTASE)	Contains protein domain (PF01678) - Metalloenzyme superfamily	264602
152	95002877 (303, 304)	Novel Protein sim. GBank gii2497952 sp P55667 Y4TM_RHISN - HYPOTHETICAL HYDROLASE/PEPTIDASE Y4TM	peptidase	264602
153	80256665 (305, 306)	Novel Protein sim. GBank gii3123021 sp Q0508 VIT1_FUNHE - VITELLOGENIN I PRECURSOR (VTG I) (CONTAINS: LIPOVITELLIN 1 (LV1); PHOSVITIN (PV); LIPOVITELLIN 2 (LV2))	UNCLASSIFIED	264593
154	82305966 (307, 308)	Novel Protein sim. GBank gii419697 pir JN0443 - transcription initiation factor sigma homolog hrdB - Streptomyces aureofaciens	Contains protein domain (PF00140) - Sigma-70 factor	264910, 264762, 264691, 264634 264605
155	20429859 (309, 310)	Novel Protein sim. GBank gii282710 pir J41739 - hypothetical protein - Escherichia coli	UNCLASSIFIED	264565
156	39564742 (311, 312)	Novel Protein sim. GBank gii3695013 (AF052586) - CIRA (Pseudomonas aeruginosa)	Contains protein domain (PF00142) - hydrolase	264691
157	10358887 (313, 314)	Novel Protein sim. GBank gii1073072 pir JC5543 - cmaU protein - Pseudomonas syringae pv. syringae	UNCLASSIFIED	264905
158	79761936 (315, 316)	Novel Protein sim. GBank gii1073072 pir JC5543 - cmaU protein - Pseudomonas syringae pv. syringae	UNCLASSIFIED	265008
159	78890376 (317, 318)	Novel Protein sim. GBank gii1173023 sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	Contains protein domain (PF00400) - VD domain, G-beta repeat	264605
160	11075119 (319, 320)	Novel Protein sim. GBank gii5304869 emb CAB46028.1  - (AL031685) dJ963K23.2 (novel protein) [Homo sapiens]	Contains protein domain (PF00327) - ribosomal prot	22278996, 264600, 264603, 35695917, 32833986, 35696423, 264636
161	80055007 (321, 322)	Novel Protein sim. GBank gii5304869 emb CAB46028.1  - (AL031685) dJ963K23.2 (novel protein) [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	264112, 264532, 22279002
162	80016371 (323, 324)	Novel Protein sim. GBank gii2661691 emb CAA15785  - (AL008204) putative protease [Streptomyces coelicolor]	UNCLASSIFIED	264639
163	11692306 (325, 326)	Novel Protein sim. GBank gii4416478 gb AAD20378  - (AF125999) transposase [Mycobacterium avium]	UNCLASSIFIED	264605, 264907, 264600
164	80077902 (327, 328)	Novel Protein sim. GBank gii76177 pir JQ0ECFT - hypothetical 38.8K protein (ftsI 5' region) - Escherichia coli	UNCLASSIFIED	264691
165	10856067 (329, 330)		UNCLASSIFIED	264605, 264486
166	80095003 (331, 332)		UNCLASSIFIED	265010
167	16395460 (333, 334)		UNCLASSIFIED	264600
168	80079362 (335, 336)			264556, 264557, 264558, 264559
169	80239561 (337, 338)			

170	79612364 (339, 340)	Novel Protein sim. GBank gi 140888 sp P27847 YIGK_ECOLI - HYPOTHETICAL 15.4 KD PROTEIN IN RECO-PL08 INTERGENIC REGION (F138)	Contains protein domain (PF01810) - LySE type translocator	264906 264595, 264604
171	95293073 (341, 342)			
172	37797007 (343, 344)	Novel Protein sim. GBank gi 4210905 gb AAD12048.1  - (AF045609) AgIG (Sinorhizobium meliloti)	Contains protein domain (PF00328) - Binding-protein-dependent transport systems inner membrane component	264769
173	57526660 (345, 346)	Novel Protein sim. GBank gi 132854 sp P02337 RL2_ECOLI - - 50S RIBOSOMAL PROTEIN L2	Contains protein domain (PF00181) - Ribosomal Proteins L2	264769
174	95293078 (347, 348)	Novel Protein sim. GBank gi 1881350 gb BAA113371  - (AB001489) PROBABLE TRANSPORT PROTEIN, SIMILAR TO ANTIBIOTIC TRANSPORT-ASSOCIATED PROTEIN ACTII IN STREPTOMYCES COELICOLOR, [Bacillus subtilis]	transport	264510, 264593, 264602, 264603, 264605, 264762, 264693
175	78756270 (349, 350)	Novel Protein sim. GBank gi 2072722 emb CAB06326  - (Z95121) manA [Mycobacterium tuberculosis]	isomerase	264565
176	80066896 (351, 352)	Novel Protein sim. GBank gi 1055198 (U40187) - similar to PIR:A41724 chicken LD (limb deformity) gene product and to formin; also P-rich region similar to collagen [Caenorhabditis elegans]	UNCLASSIFIED	264907, 264910, 264681, 264558
177	86684832 (353, 354)	Novel Protein sim. GBank gi 2328739 emb CAB10952  - (Z98268) hypothetical protein RV1695 [Mycobacterium tuberculosis]	Contains protein domain (PF01513) - Domain of unknown function	264768, 60424179, 264687, 264688, 264769, 29331826, 60432289, 18108376, 264689, 18108387, 32833986, 22278996, 265020, 264800, 264801, 264802, 264803, 264804, 264805, 264835, 264762, 264638, 264906, 264564, 264637, 264638, 264486, 60433358, 264766
178	79559526 (355, 356)	Novel Protein sim. GBank gi 1906596 (U81788) - kinesin-73 [Drosophila melanogaster]	struct	264693, 33657109, 264635
179	20263112 (357, 358)	Novel Protein sim. GBank gi 1169367 sp P45256 DNAB_HAEIN - REPLICATIVE DNA HELICASE	UNCLASSIFIED	264563
180	80488958 (359, 360)	Novel Protein sim. GBank gi 3170615 (AF059485) - DOC4 [Mus musculus]	helicase	264769
181	79585389 (361, 362)		UNCLASSIFIED	21806767, 264635, 264639, 18108384
182	80577899 (363, 364)		UNCLASSIFIED	264259, 35696052, 56182435, 264511, 265018, 33657109, 264555, 264568
183	11614017 (365, 366)	Novel Protein sim. GBank gi 1076627 pir S54172 - Inorganic pyrophosphatase (EC 3.6.1.1) - common tobacco	UNCLASSIFIED	264690
184	10174167 (367, 368)	Novel Protein sim. GBank gi 4371280 gb AAD18138  - (AC008260) hypothetical protein [Arabidopsis thaliana]	UNCLASSIFIED	264510

185	21660822 (369, 370)	Novel Protein sim. GBank gij3006178[emb]CAA18398.1] - (AL022304) putative mma transport regulator [Schizosaccharomyces pombe]		UNCLASSIFIED	264604
186	80070329 (371, 372)	Novel Protein sim. GBank gij2829802[sp]P94408[YCLF_BACSU - HYPOTHETICAL 53.3 KD PROTEIN IN SFP-GERKA INTERGENIC REGION]		transport	264595
187	80186611 (373, 374)	Novel Protein sim. GBank gij3150260[emb]CAA19179] - (AL023634) cyclin [Schizosaccharomyces pombe]		UNCLASSIFIED	264369
188	20464942 (375, 376)	Novel Protein sim. GBank gij2145853[pri]S72938 - htlx protein - Mycobacterium leprae		kinase	264605
189	82338215 (377, 378)	Novel Protein sim. GBank gij1881244[db]BAA19271] - (AB001488) SIMILAR TO PYRUVATE OXIDASE AND ACETOLACTATE SYNTHASE. [Bacillus subtilis]		UNCLASSIFIED	35898052, 264602, 264605, 264762, 264689, 35898917, 18108370, 18108372, 264638, 264595
190	80086821 (379, 380)	Novel Protein sim. GBank gij1705461[sp]P53656[BIOA_ERWHE - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7.8-DIAMINO-PELAGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)]	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	- synthase	264563
191	88095012 (381, 382)	Novel Protein sim. GBank gij120226[sp]P28725[FKBP_STRCH - FKBP-type peptidyl-prolyl cis-trans isomerases (PIPIASE) (ROTAMASE)]	Contains protein domain (PF00254) - FKBP-type peptidyl-prolyl cis-trans isomerases	isomerase	284508, 264604, 264605, 264769, 264555
192	16333379 (383, 384)	Novel Protein sim. GBank gij4980892[gb]AAD35474.1[AE00171] - (AE001718) ABC transporter, ATP-binding protein [Thermotoga maritima]			264567
193	78910127 (385, 386)	Novel Protein sim. GBank gij1705461[sp]P53656[BIOA_ERWHE - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7.8-DIAMINO-PELAGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)]			284608, 264693
194	20464943 (387, 388)	Novel Protein sim. GBank gij1705461[sp]P53656[BIOA_ERWHE - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7.8-DIAMINO-PELAGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)]			264605
195	13518389 (389, 390)	Novel Protein sim. GBank gij1705461[sp]P53656[BIOA_ERWHE - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7.8-DIAMINO-PELAGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)]		transport	264635
196	95005569 (391, 392)	Novel Protein sim. GBank gij1705461[sp]P53656[BIOA_ERWHE - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7.8-DIAMINO-PELAGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)]	Contains protein domain (PF00202) - Aminotransferases daas-III pyridoxal phosphate	- gaba	264600, 264689, 264638
197	80248665 (393, 394)	Novel Protein sim. GBank gij3122305[sp]Q27778[K6PF_SCHMA - 6-PHOSPHOFRUCTOKINASE (PHOSPHOFRUCTOKINASE)]	Contains protein domain (PF00365) - Phosphofructokinase	kinase	264602, 264682, 264692, 18108374
198	79163635 (395, 396)	Novel Protein sim. GBank gij1781203[emb]CAB06110] - (Z83859) gnd [Mycobacterium tuberculosis]			264636
199	78950715 (397, 398)	Novel Protein sim. GBank gij1781203[emb]CAB06110] - (Z83859) gnd [Mycobacterium tuberculosis]	Contains protein domain (PF00383) - 6-phosphogluconate dehydrogenases		265008
200	79413648 (399, 400)	Novel Protein sim. GBank gij2642222 [AF030885] - telomere-associated recQ-like helicase [Ustilago maydis]		UNCLASSIFIED	264595, 264596
201	86945924 (401, 402)	Novel Protein sim. GBank gij2894378[emb]CAA74811.1] - (Y14573) ring finger protein [Hordeum vulgare]		UNCLASSIFIED	29331826, 285007, 264512, 33657402, 264596, 285017, 18108351, 284682, 264683, 264767, 264628, 55810794, 264634, 264635, 56182323, 60432113, 22279000

202	78568046 (403, 404)	Novel Protein sim. GBank gi 231772 sp P30598 CHST1 - CHITIN SYNTHASE 1 (CHITIN-UDP ACETYL-GLUCOSAMINYL TRANSFERASE 1)	Contains protein domain (PF01844) - Chitin synthase	- synthase	264600
203	79843927 (405, 406)	Novel Protein sim. GBank gi 1504042 dbj BAA13220  - (D88984) similar to yeast adenylate cyclase (S56776) [Homo sapiens]			22278995, 29331822, 29331825, 29331827, 264906, 21908754, 264683, 21908766, 21908769, 33586423, 264556
204	79855185 (407, 408)			UNCLASSIFIED	264909
205	10080583 (409, 410)	Novel Protein sim. GBank gi 2633808 emb CAB13310  - (Z59111) similar to hypothetical proteins [Bacillus subtilis]		transport	264909
206	8758473 (411, 412)			UNCLASSIFIED	264604
207	20754522 (413, 414)	Novel Protein sim. GBank gi 2134381 pir S60678 - polybromo 1 protein - chicken		UNCLASSIFIED	264556
208	20289281 (415, 416)				264605
209	80071069 (417, 418)	Novel Protein sim. GBank gi 2501040 sp O5814 SYP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LIGASE) (PRORS)			264605, 264689
210	80168800 (419, 420)				264905, 264907, 264909, 264766, 264687, 264691, 264629, 18108374, 264638
211	80034539 (421, 422)				263978
212	82442474 (423, 424)	Novel Protein sim. GBank gi 5031809 ref NP_005536.1 pISLR - immunoglobulin superfamily containing leucine-rich repeat		UNCLASSIFIED	264508, 264905, 264906, 264907, 264908, 264600, 264762, 264534, 264632, 264634, 264635, 264639, 264486
213	80248562 (425, 426)	Novel Protein sim. GBank gi 3122359 sp O33123 LEU2_MYCLE - 3- ISOPROPYLMALATE DEHYDRATASE LARGE SUBUNIT (ISOPROPYLMALATE ISOMERASE) (ALPHA-IPM ISOMERASE) (IPMI)	Contains protein domain (PF00330) - isomerase Aconitase family (aconitate hydratase)		22278996, 264508, 264600, 264602, 264603, 264605, 33657023, 264565, 264486
214	80079381 (427, 428)	Novel Protein sim. GBank gi 116238 sp P19421 CH60_COXBU - 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN B)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	- eph	264600, 264693
215	14873283 (429, 430)				
216	80177716 (431, 432)	Novel Protein sim. GBank gi 2506924 sp P48754 VP41_HUMAN - VACUOLAR ASSEMBLY PROTEIN VPS41 HOMOLOG (S53)	Contains protein domain (PF00096) - dna_ma_bind Zinc finger, C2H2 type	UNCLASSIFIED	264629 264448
217	79603634 (433, 434)				264508
218	80258475 (435, 436)	Novel Protein sim. GBank gi 173288 sp P38108 RSEA_ECOLI - SIGMA-E FACTOR NEGATIVE REGULATORY PROTEIN		mapolymerase	264594
219	20438787 (437, 438)	Novel Protein sim. GBank gi 1781097 emb CAB06231  - (Z63864) gltB [Mycobacterium tuberculosis]		synthase	264604
220	13489572 (439, 440)	Novel Protein sim. GBank gi 2384703 AF052427  - unknown [Trypanosoma cruzi]		nucleaseinhib	264689
221	11287498 (441, 442)	Novel Protein sim. GBank gi 4587313 dbj BAA76709.1  - (AB025248) alpha-1,2-mannosidase [Bacillus sp. M-90]		UNCLASSIFIED	264555



222	79862802 (443, 444)	Novel Protein sim. GBank gij187726[embjCAB07049] - (Z92770) hypothetical protein Rv0143c [Mycobacterium tuberculosis]		UNCLASSIFIED	264605, 264769, 35698423
223	83053869 (445, 446)			UNCLASSIFIED	264906, 264907, 264603
224	79557820 (447, 448)			UNCLASSIFIED	264684, 264693
225	79559541 (449, 450)	Novel Protein sim. GBank gij2274851[dbjBAA21515] - (D64159) 3-7 gene product [Homo sapiens]		UNCLASSIFIED	264692
226	79172397 (451, 452)	Novel Protein sim. GBank gij868245 (U29488) - C56C10.7 gene product [Caenorhabditis elegans]		UNCLASSIFIED	22278998, 264112, 33657023, 263981
227	811777196 (453, 454)			UNCLASSIFIED	35695917, 264636, 264907
228	79872885 (455, 456)				264768, 264907, 264908, 264692, 264593, 264639
229	79838268 (457, 458)				264906, 264910
230	11013209 (459, 460)			UNCLASSIFIED	264631
231	20822207 (461, 462)	Novel Protein sim. GBank gij1835114[embjCAA71733] - (Y10744) homoserine O-acetyltransferase [Leptospira meyeri]			264906, 264600, 264603, 264692
232	80055035 (463, 464)	Novel Protein sim. GBank gij2642340 (AF032970) - imidazole propionate hydrolase [Pseudomonas putida]	Contains protein domain (PF00449) - Urease	UNCLASSIFIED	264600, 264603, 264605, 264687, 264769
233	80063054 (465, 466)	Novel Protein sim. GBank gij3510503 (AF030881) - polypeptide [Fugu rubripes]		UNCLASSIFIED	264604
234	7523598 (467, 468)	Novel Protein sim. GBank gij2104608[embjCAB088005] - (Z95398) Peka [Mycobacterium leprae]		UNCLASSIFIED	264369
235	80203671 (469, 470)			UNCLASSIFIED	264108
236	78940001 (471, 472)			UNCLASSIFIED	264905
237	11755273 (473, 474)				264681
238	79461401 (475, 476)	Novel Protein sim. GBank gij2495617[spjQ57252]YDIJ_HAEIN - HYPOTHETICAL PROTEIN H1163	Contains protein domain (PF00037) - 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains.	UNCLASSIFIED	264639
239	82435180 (477, 478)	Novel Protein sim. GBank gij3183458[spjP75798]YLIA_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YLIA			264906, 265010, 264603, 264762, 264682, 264636, 264638, 264488
240	21635575 (479, 480)	Novel Protein sim. GBank gij281503 predicted using GeneFinder; similar to collagen; cDNA EST EMBL:D65450 comes from this gene; cDNA EST EMBL:D68888 comes from this gene [Caenorhabditis elegans]	transport		264259, 264769
241	80377307 (481, 482)	Novel Protein sim. GBank gij2624302[embjCAB04111] - (Z81503) predicted using GeneFinder; similar to collagen; cDNA EST EMBL:D65450 comes from this gene; cDNA EST EMBL:D68888 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	264908, 264909, 264764, 264639
242	82146454 (483, 484)			UNCLASSIFIED	264489, 264907, 264908, 264511, 264760, 264764, 264692, 264635, 264637
243	79633207 (485, 486)	Novel Protein sim. GBank gij2624302[embjCAA15375] - (AL008987) ald [Mycobacterium tuberculosis]		UNCLASSIFIED	264906
244	80246682 (487, 488)	Novel Protein sim. GBank gij2920625 (AF044499) - vgrE protein [Escherichia coli]		UNCLASSIFIED	264600, 264602, 264605, 264769, 264689
245	79863543 (489, 490)	Novel Protein sim. GBank gij5420387[embjCAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264907, 264758
246	79162929 (491, 492)		Contains protein domain (PF01108) - NRU-like domain		264637, 18108381, 18108387, 264565

247	79873185 (493, 494)	Novel Protein sim. GBank gll1839006[emb]CA806646] - (Z85982) arpB [Mycobacterium tuberculosis]		kinase	284909, 284691, 35698423, 18108387
248	80489983 (495, 496)	Novel Protein sim. GBank gll1168574[sp]P42464[ATPB_CORGL - ATP SYNTHASE BETA CHAIN		synthase	35698286, 284907, 284511, 284602, 284768, 284688, 285021, 35695855, 18108385
249	79784848 (497, 498)			UNCLASSIFIED	284907, 284910, 285011, 284762, 284638
250	79619980 (499, 500)				21906768, 284692
251	84359489 (501, 502)			UNCLASSIFIED	52645138, 29331822, 29331824, 52644045, 285018, 21906765, 21906768, 285020, 27486261, 27486265, 35695763, 18108376, 284536, 284559, 284565
252	79737756 (503, 504)	Novel Protein sim. GBank gll3327166[dbj]BAA31651] - (AB014576) KIAA0878 protein [Homo sapiens]			284685, 284687, 284632
253	20443124 (505, 506)	Novel Protein sim. GBank gll3036880[emb]CAA18513] - (AL022374) putative ATP-dependent DNA helicase [Streptomyces coelicolor]		helicase	284604
254	80027421 (507, 508)	Novel Protein sim. GBank gll3915488[sp]O34961]YJMB_BACSU - HYPOTHETICAL SYMPORTER IN COTI-RAPA INTERGENIC REGION		UNCLASSIFIED	284508, 284906, 284602, 284687, 285021, 284488
255	11398315 (509, 510)	Novel Protein sim. GBank gll1665720[dbj]BAA04134] - (D17312) dietheral toxin [Bacillus cereus]		UNCLASSIFIED	284593
256	80028158 (511, 512)	Novel Protein sim. GBank gll465787[sp]P42423]YL31_CAEEL - HYPOTHETICAL 86.0 KD PROTEIN F4489.1 IN CHROMOSOME II	Contains protein domain (PF00326) - Prolyl oligopeptidase family	peptidase	284602, 284692
257	20289282 (513, 514)	Novel Protein sim. GBank gll1172038[sp]P42315]SCOA_BACSU - PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE SUBUNIT A (SUCCINYL COA:3-OXOACID COA-TRANSFERASE) (OXCT A)	Contains protein domain (PF01144) - Coenzyme A transferase	transferase	284605
258	20459484 (515, 516)	Novel Protein sim. GBank gll3127838[emb]CAA18902] - (AL023496) hypothetical protein [Streptomyces coelicolor]		UNCLASSIFIED	284604
259	79910152 (517, 518)			collagen	284681, 284686, 284692
260	20379437 (519, 520)			UNCLASSIFIED	284692, 284556
261	20285883 (521, 522)	Novel Protein sim. GBank gll123761[sp]P24221]HUTH_S.TRGR - HISTIDINE AMMONIA-LYASE (HISTIDASE)	Contains protein domain (PF00221) - Phenylalanine and histidine ammonia lyases	UNCLASSIFIED	284600
262	80189317 (523, 524)			UNCLASSIFIED	285017, 284369
263	88085045 (525, 526)	Novel Protein sim. GBank gll3924708[emb]CAA84646] - (Z35587) Weak similarity with sea squirt nidogen precursor protein (blast score 71); cDNA EST EMBL: T02069 comes from this gene; cDNA EST EMBL: D78135 comes from this gene; cDNA EST EMBL: D73147 comes from this gene; cDNA EST EMBL: ...		UNCLASSIFIED	284488, 284905, 284906, 284907, 284908, 284909, 284512, 284910, 284758, 284596, 284604, 285018, 284605, 284760, 18108351, 284783, 284784, 284288, 284768, 284768, 284769, 284691, 284692, 284693, 284628, 284634, 284635, 284555, 284638, 284639
264	87370826 (527, 528)	Novel Protein sim. GBank gll3043734[dbj]BAA25531] - (AB011177) KIAA0805 protein [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	protease	284259, 284908, 21906754, 285018, 285019, 285020

265	95355646 (529, 530)	Novel Protein sim. GBank gi 4589624 d JBAA76834.1  - (AB023207) KIAA0990 protein [Homo sapiens]	kinase	264488, 35696286, 29331624, 56182181, 35696052, 264508, 264905, 264806, 284907, 66712502, 264908, 264909, 264511, 264512, 264910, 264592, 264595, 264768, 264596, 55811386, 264600, 265017, 264603, 284604, 264605, 264780, 18108351, 264762, 284681, 264784, 264288, 264766, 264768, 264769, 21906765, 21906767, 21906769, 265020, 264691, 33657023, 33657109, 33657182, 264628, 35696423, 35695855, 264630, 264631, 264632, 264634, 264635, 284636, 264555, 264638, 83373044, 56528488, 87188518, 264564, 264568, 264488, 264600
266	79580075 (531, 532)		UNCLASSIFIED	264600
267	11362222 (533, 534)		UNCLASSIFIED	264828
268	79809566 (535, 536)		UNCLASSIFIED	264687, 264769, 264689
269	80025810 (537, 538)		UNCLASSIFIED	264602
270	84361144 (539, 540)	Novel Protein sim. GBank gi 4507367 ref NP_003182.1 pTARS - Ihteenyl-IRNA synthetase	UNCLASSIFIED	264693
271	79552301 (541, 542)	Novel Protein sim. GBank gi 4980738 gb AA035331.1 AE00170 - (AE001707) glucose-1-phosphate adenylyltransferase [Thermotoga maritima]	UNCLASSIFIED	264909, 264693
272	9874778 (543, 544)		synthase	264908
273	12840594 (545, 546)	Novel Protein sim. GBank gi 1168224 sp P44569 5NTD_HAEIN - PROBABLE 5'-NUCLEOTIDASE PRECURSOR	UNCLASSIFIED	264688
274	39524246 (547, 548)	Novel Protein sim. GBank gi 3253159 (AF005355) - translation initiation factor eIF2C [Oryctolagus cuniculus]	UNCLASSIFIED	264564
275	82787041 (549, 550)	Novel Protein sim. GBank gi 1349201 sp P21997 SSGP_VOLCA - SULFATED SURFACE GLYCOPROTEIN 185 (SSG 185)	UNCLASSIFIED	264807, 264908, 264809, 264768, 264768, 264691, 264632, 264638
276	86671073 (551, 552)	Novel Protein sim. GBank gi 128021 sp P20964 OBG_BACSU - SPO08-ASSOCIATED GTP-BINDING PROTEIN		265008, 60432229
277	80079735 (553, 554)		ribosomal prot	264600, 18108387
278	12866947 (555, 556)	Novel Protein sim. GBank gi 79839 p J S03812 - uvrB protein - Micrococcus luteus	UNCLASSIFIED	264689
279	95292719 (557, 558)		nuclease	264508, 264604, 21906764, 264638, 264557, 264404
280	5803617 (559, 560)			264259
281	80249599 (561, 562)	Novel Protein sim. GBank gi 3123160 sp Q18964 YLN2_CAEEL - HYPOTHETICAL 46.2 KD TRP-ASP REPEATS CONTAINING PROTEIN D2013.2 IN CHROMOSOME II		18108392, 264634, 264555, 264556, 264557, 264558
282	18598882 (563, 564)		UNCLASSIFIED	265018
283	20814211 (565, 566)		UNCLASSIFIED	264555

284	91212160 (567, 568)	Novel Protein sim. GBank gij2429094 (U58632) - acetyl xylan esterase; AxeA [Thermotoga neapolitana]	Contains protein domain (PF00300) - Phosphoglycerate mutase family	UNCLASSIFIED	35696052, 29331828, 264508, 264905, 264600, 264602, 264605, 264682, 264764, 56181562, 21806764, 18108376, 264638, 264559, 18108387
285	8757840 (569, 570)	Novel Protein sim. GBank gij2072674 (embj[CAB08305] - (Z95120) rhlE [Mycobacterium tuberculosis]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	UNCLASSIFIED	264603
286	80503235 (571, 572)	Novel Protein sim. GBank gij1870009 (embj[CAB06860] - (Z92539) hypothetical protein Rv1019 [Mycobacterium tuberculosis]	UNCLASSIFIED	ATPase associated	264638
287	12745521 (573, 574)	Novel Protein sim. GBank gij785323 (bbsj[57678] - (S74439) silk fibroin heavy chain (C-terminal) [Bombyx mori-silkworms, Peptide Partial, 633 aa] [Bombyx mori]	UNCLASSIFIED	collagen	264689
288	20756502 (575, 576)	Novel Protein sim. GBank gij1870009 (embj[CAB06860] - (Z92539) hypothetical protein Rv1019 [Mycobacterium tuberculosis]	UNCLASSIFIED	collagen	264557
289	80043804 (577, 578)	Novel Protein sim. GBank gij1870009 (embj[CAB06860] - (Z92539) hypothetical protein Rv1019 [Mycobacterium tuberculosis]	Contains protein domain (PF00440) - Bacterial regulatory proteins, tetR family	UNCLASSIFIED	264593, 264600
290	80430175 (579, 580)	Novel Protein sim. GBank gij2506664 (spj[P40120]YDCG_ECOLI - 59.4 PROTEIN IN TRG-RIML INTERGENIC REGION PRECURSOR	UNCLASSIFIED	UNCLASSIFIED	264768
291	20747431 (581, 582)	Novel Protein sim. GBank gij1716065 (spj[P53528]UVRD_MYCLE - PUTATIVE DNA HELICASE II HOMOLOG	UNCLASSIFIED	UNCLASSIFIED	264601
292	80052555 (583, 584)	Novel Protein sim. GBank gij1716065 (spj[P53528]UVRD_MYCLE - PUTATIVE DNA HELICASE II HOMOLOG	UNCLASSIFIED	UNCLASSIFIED	264605
293	80062518 (585, 586)	Novel Protein sim. GBank gij1716065 (spj[P53528]UVRD_MYCLE - PUTATIVE DNA HELICASE II HOMOLOG	UNCLASSIFIED	UNCLASSIFIED	264605
294	78830303 (587, 588)	Novel Protein sim. GBank gij117422 (spj[P10040]ICRB_DROME - CRUMBS PROTEIN PRECURSOR (95F)	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	264909, 264605, 264687, 264689, 264692
295	79444180 (589, 590)	Novel Protein sim. GBank gij1181819 (dbj[BAA11565] - (D82364) a variant of TSC-22 [Gallus gallus]	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	35696052, 264906, 265011, 264628, 55811576
296	79607076 (591, 592)	Novel Protein sim. GBank gij3649789 (dbj[BAA33403] - (AD012226) SecA [Vibrio alginolyticus]	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	52644507, 29331822, 264592, 265020, 264639
297	78631297 (593, 594)	Novel Protein sim. GBank gij5689967 (embj[CAB52004.1] - (AL109863) putative membrane protein [Streptomyces coelicolor A3(2)]	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	264639
298	80418898 (595, 596)	Novel Protein sim. GBank gij5689967 (embj[CAB52004.1] - (AL109863) putative membrane protein [Streptomyces coelicolor A3(2)]	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	264508
			Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	264905, 264687, 264638
			Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	264905, 264681, 264639, 264766

299	95293298 (597, 598)	Novel Protein sim. GBank gll220637[gbjBA401477] - (O10627) zinc finger protein [Mus musculus]	Contains protein domain (PF00098) - Zinc finger, C2H2 type		264488, 263994, 56994075, 22278997, 22278998, 22278999, 20281099, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 264905, 264906, 264907, 264908, 52644045, 264909, 264511, 265008, 264910, 264595, 264596, 264758, 33657084, 87168559, 265018, 265019, 264764, 264288, 264766, 264687, 56181562, 264769, 21906765, 21906768, 21906769, 33657023, 264692, 33657109, 27486281, 18108370, 264628, 264629, 55811576, 35695855, 264631, 264634, 264635, 264638, 264639, 83373044, 18108387, 87168518, 22279000, 22279002, 264565, 264568, 264567
300	20711340 (599, 600)	Novel Protein sim. GBank gll145922 (M20951) - iron	UNCLASSIFIED		264602
301	13511332 (601, 602)	dedicate transport protein precursor [Escherichia coli]	transport		264687
302	9875260 (603, 604)	Novel Protein sim. GBank gll174661[spP44594]GTG_HAEIN - QUEUINE TRNA-RIBOSYLTRANSFERASE (TRNA-GUANINE TRANSGLYCOSYLASE) (GUANINE INSERTION ENZYME)			264808
303	79574895 (605, 606)	Novel Protein sim. GBank gll67985[prj]HUNVAV - helicase			264689
304	20711344 (607, 608)	(EC 3.6.1.-) - Autographa californica nuclear polyhedrosis virus	helicase		264602
305	80412520 (609, 610)	Novel Protein sim. GBank gll728667[spP40602]APG_ARATH - ANTER-SPECIFIC PROLINE-RICH PROTEIN APG PRECURSOR			264763
306	8515878 (611, 612)	Novel Protein sim. GBank gll1657554[gbjAAB18082.1] - (U73857) hypothetical protein [Escherichia coli]	UNCLASSIFIED		263878
307	80222801 (613, 614)	Novel Protein sim. GBank gll1710612[spQ10793]RNH2_MYCTU - PROBABLE RIBONUCLEASE HII (RNASE HII)	UNCLASSIFIED		265010, 21906768, 265020, 18108374, 263977
308	80084305 (615, 616)	Novel Protein sim. GBank gll5420387[embjCAB46678.1] - (AJ243459) proteophosphoglycan [Leishmania major]	Contains protein domain (PF01351) - Ribonuclease HII		264910, 264600, 264605, 264687, 264689, 264638, 18108387
309	80504136 (617, 618)	Novel Protein sim. GBank gll1744522 (U34957) - phosphoribosylaminoimidazole succinocarboxamide synthase [Mycobacterium tuberculosis]			264769
310	80053816 (619, 620)				264603
311	11090659 (621, 622)				264602
312	80054347 (623, 624)				
313	80046168 (625, 626)		UNCLASSIFIED		264566
					264603, 264567

314	87845112 (627, 628)	Novel Protein sim. GBank gl 3651583 (AF092175) - Ikaros [Dario reio]	Contains protein domain (PF00320) - GATA zinc finger	dna_ma_bind	264259, 60432289, 26331828, 264905, 264908, 264908, 264909, 265008, 264910, 60432229, 33657402, 60433438, 33109954, 265011, 265017, 264603, 265018, 264288, 264766, 264692, 35695763, 264628, 264629, 264639, 60170394, 22279002, 264568
315	82356091 (629, 630)	Novel Protein sim. GBank gl 1652620(jb BAA17540) - (D90907) pyridine nucleotide transhydrogenase beta subunit [Synectocystis sp.]			264508, 264600, 264762, 264687, 264768, 52644229, 264769, 264689, 264635, 264636, 264638, 264488
316	78911071 (631, 632)	Novel Protein sim. GBank		UNCLASSIFIED	264693
317	20466944 (633, 634)	Novel Protein sim. GBank gl 18244sp P24176DAPE_ECOLI - SUCCINYL-DIAMINOPIMELATE DESUCCINYLASE (SDAP)		UNCLASSIFIED	264605
318	94141836 (635, 636)	Novel Protein sim. GBank gl 4680229(jb AAD27583.1 AF11827) - (AF118274) DNB-5 [Homo sapiens]	Contains protein domain (PF00526) Dictyostellum (slime mold) repeats	transport	264908, 264909, 264910, 264593, 264594, 264760, 264288, 264768, 264769, 21906769, 264691, 264693, 264628, 63274791, 264635, 264638, 264638, 83373044, 22279002, 264588
319	17289360 (637, 638)	Novel Protein sim. GBank gl 1149689(jemb CAA60220) - (X86499) rbsC [Clostridium perfringens]		transport	265018
320	13327675 (639, 640)	Novel Protein sim. GBank gl 281103sp Q05314(GLGC_MYCTU - GLUCOSE-1-PHOSPHATE ADENYLTRANSFERASE (ADP-GLUCOSE SYNTHASE) (ADP-GLUCOSE PYROPHOSPHORYLASE)		synthase	264667
321	94134387 (641, 642)	Novel Protein sim. GBank gl 1680716 (U68234) - all-trans-retinoic acid 4-hydroxylase [Dario reio]		cyto450	264509, 264906, 264907, 264908, 265009, 264596, 264764, 264628, 264634, 264635, 264638, 264639, 83373044, 264567
322	66489053 (643, 644)	Novel Protein sim. GBank gl 1160355 (U33058) - UNC-89 [Caenorhabditis elegans]		UNCLASSIFIED	55811150, 264691, 60431528, 55810764
323	94553725 (645, 646)			UNCLASSIFIED	264488, 265009, 264593, 264628, 264635
324	78174383 (647, 648)			UNCLASSIFIED	264687
325	79862691 (649, 650)			UNCLASSIFIED	264693
326	28774974 (651, 652)			UNCLASSIFIED	264288, 18108385
327	79776267 (653, 654)	Novel Protein sim. GBank gl 451544 (U04267) - proline-rich cell wall protein [Gossypium barbadense]			264488, 264905, 264509, 264910
328	80253202 (655, 656)			UNCLASSIFIED	264592
329	10173821 (657, 658)			UNCLASSIFIED	264510
330	86597767 (659, 660)	Novel Protein sim. GBank gl 4191359 (AF087825) - claudin-7 [Mus musculus]		UNCLASSIFIED	264258, 264908
331	79754888 (661, 662)	Novel Protein sim. GBank gl 80741(pir S20912) - regulatory protein whiB - Streptomyces coelicolor		transcriptfactor	264910, 264687, 264688, 264638, 264567
332	80071440 (663, 664)	Novel Protein sim. GBank gl 114049sp P19480(AHPF_SALTY - ALKYL HYDROPEROXIDE REDUCTASE SUBUNIT F (ALKYL HYDROPEROXIDE REDUCTASE F52A PROTEIN)		reductase	35696423, 264636, 264638, 264555
333	13009555 (665, 666)				264687

334	80230771 (667, 668)	Novel Protein sim. GBank gij3222228[pir][S32227 - glutamate dehydrogenase (NADP+)] (EC 1.4.1.4) - Corynebacterium glutamicum	Contains protein domain (PF00208) - Glutamate/Leucine/Phenylalanine/Valine dehydrogenase	dehydrogenase	264905, 264800, 264804, 264486
335	80057026 (668, 670)	Novel Protein sim. GBank gij2193838[emb][CAB09602] - (Z96800) gipQ2 [Mycobacterium tuberculosis]		esterase	264907, 264803, 264693, 18108374, 264636, 18108387
336	80414319 (671, 672)			UNCLASSIFIED	265009, 264766, 264686
337	11090829 (673, 674)				264602
338	95413134 (675, 676)	Novel Protein sim. GBank gij5454074[refNP_006303.1]pSMRT - silencing mediator for retinoid and thyroid hormone receptors	Contains protein domain (PF00249) - Myb-like DNA-binding domain	nuc_d_rept	264569, 18108397, 22278998, 29331822, 20281099, 29331824, 56182161, 68714117, 29331825, 35696052, 29331826, 264508, 264509, 264905, 264806, 264907, 264908, 264909, 265008, 265006, 264910, 265009, 264758, 55812038, 65274444, 265011, 87168559, 265017, 265018, 265019, 264760, 55811150, 264681, 264762, 18108351, 264682, 264764, 264766, 264689, 55811957, 264768, 52844228, 264689, 55811957, 35695917, 264692, 264693, 264628, 18108370, 18108374, 55811576, 35699423, 35695855, 264635, 264555, 264636, 264558, 264637, 264557, 18108380, 264638, 264558, 264639, 18108381, 83373044, 18108385, 87168518, 60432113
339	11398513 (677, 678)	Novel Protein sim. GBank gij4001713[dbj]BA435087.1] - (AB015878) DnaK [Pseudomonas gingivalis]		eph	264593
340	80504148 (679, 680)	Novel Protein sim. GBank gij2842899[spQ92353]UBPC_SCHPO - PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C8G9.08 (UBIQUITIN THIOLESTERASE) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE) (DEUBIQUITINATING ENZYME)		ubiquitin	264905, 265019, 264769, 18108374
341	11075198 (681, 682)	Novel Protein sim. GBank gij2688560 (AE001166) - conserved hypothetical protein [Borrelia burgdorferi]	Contains protein domain (PF00290) - Tryptophan synthase alpha chain	isomerase	264605
342	80054186 (683, 684)	Novel Protein sim. GBank gij1684738[emb][CAA70601] - (Y09452) Yed J hypothetical protein [Pseudomonas syringae]			264603, 264604
343	20468782 (685, 686)				
344	80428670 (687, 688)	Novel Protein sim. GBank gij2117275[emb][CAB09104] - (Z95618) hypothetical protein Rv0807 [Mycobacterium tuberculosis]		UNCLASSIFIED	264605, 264605, 264788, 18108370, 18108374, 35695855
345	80258853 (689, 690)	Novel Protein sim. GBank gij3023317[spQ48935]APHA_MYCRA - ACETYL POLYAMINE AMINOHYDROLASE		histone	264593
346	78831058 (691, 692)	Novel Protein sim. GBank gij4239787[emb][CAA75437] - (Y15168) NADP-glutamate dehydrogenase [Pseudomonas aeruginosa]	Contains protein domain (PF00208) - Glutamate/Leucine/Phenylalanine/Valine dehydrogenase	dehydrogenase	264905

347	79158185 (693, 694)	Novel Protein sim. GBank gi 731675 sp P38795 YH4_YEAST - HYPOTHETICAL 80.7 KD PROTEIN IN ERG7-NMD2 INTERGENIC REGION protein - Escherichia coli		UNCLASSIFIED	255006, 265008, 265010, 265018, 263967, 263981
348	80020208 (695, 696)	Novel Protein sim. GBank gi 1073610 pf S47672 - ugpB		transport	264602, 18108351, 18108387
349	17282112 (697, 698)				265007
350	80502370 (699, 700)	Novel Protein sim. GBank gi 3261599 mb CAB009171 - (Z77137) hypothetical protein Rv1277 [Mycobacterium tuberculosis]		nuclease	265009, 264769, 264689, 18108370
351	80501805 (701, 702)	Novel Protein sim. GBank gi 2959367 mb CAA179211 - (AL022117) hypothetical protein [Schizosaccharomyces pombe]		glycoprotein	264769, 264905, 264908
352	11811585 (703, 704)	Novel Protein sim. GBank gi 416302 gb AAD203071 - (AF105718) copia-type pol polyprotein [Zea mays]		protease	264595
353	80061653 (705, 706)	Novel Protein sim. GBank gi 1174887 sp P42873 URE1_STAXY - UREASE ALPHA SUBUNIT [UREA AMIDOHYDROLASE]	Contains protein domain (PF00449) - Urease	UNCLASSIFIED	264604
354	56626130 (707, 708)			UNCLASSIFIED	264628
355	80046344 (709, 710)	Novel Protein sim. GBank gi 115157 sp P16574 BVGA_BORPE - VIRULENCE FACTORS PUTATIVE POSITIVE TRANSCRIPTION REGULATOR BVGA	Contains protein domain (PF00072) - Response regulator receiver domain	UNCLASSIFIED	264909, 264595, 264683, 22279002 264909, 264591, 264592
356	80043835 (711, 712)	Novel Protein sim. GBank gi 497637 (J03939) - cytochrome oxidase d subunit I [Escherichia coli]		oxidase	264605
357	80070566 (713, 714)	Novel Protein sim. GBank gi 2280990 (AF006000) - Brg1 [Bordetella pertussis]		UNCLASSIFIED	264768
358	37032756 (715, 716)			UNCLASSIFIED	264604, 264769
359	80501488 (717, 718)	Novel Protein sim. GBank gi 3510639 (AF049344) - UDP- GalNAc:polypeptide N-acetylglucosaminyltransferase T5 [Rattus norvegicus]		UNCLASSIFIED	264694
360	80026748 (719, 720)			transferase	22278996, 264259, 26331822, 29331824, 264605, 55811857, 265022
361	80584075 (721, 722)				
362	13089485 (723, 724)	Novel Protein sim. GBank gi 113764 sp P25718 AMY1_ECOLI - ALPHA-AMYLASE PRECURSOR (1.4-ALPHA-D-GLUCAN GLUCANOHYDROLASE)		amylase	264688
363	70750145 (725, 726)				
364	82443563 (727, 728)	Novel Protein sim. GBank gi 2829818 sp P95171 NUOK_MYCTU - NADH DEHYDROGENASE I CHAIN K (NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 11) [NUO11]	Contains protein domain (PF00420) - NADH-ubiquinone:plastoquinone oxidoreductase chain 4L	dehydrogenase	264568 264769, 264602, 264604, 264508, 264762, 264638, 264486



365	88040288 (729, 730)	Novel Protein sim. GBank gi 4929268 gb AAD33924.1  - (AF144237) LOMP protein [Homo sapiens]	Contains protein domain (PF00412) - LIM domain containing proteins	264488, 21906766, 21908767, 55811576, 21906769, 26148629, 22278995, 22278996, 265020, 265022, 264634, 264991, 264992, 33857023, 33857402, 264693, 264639, 264594, 29331824, 264758, 18108385, 29331827, 8718659, 265016, 22279000, 265019, 264482, 264761, 264681, 18108351, 265017, 264757
366	81821838 (731, 732)			
367	95357471 (733, 734)	Novel Protein sim. GBank gi 4503843 ref NP_003908.1 pG2AD - UNKNOWN	Contains protein domain (PF01602) - Adaplin N terminal region	60424179, 63274572, 56182575, 22278994, 56994075, 22278998, 264259, 29331822, 29331824, 56182181, 60424259, 68714117, 29331825, 60432289, 29331826, 29331827, 29331828, 264905, 264828, 56182435, 265006, 264512, 265008, 264591, 55812038, 55811366, 265010, 87186599, 265017, 265018, 264604, 265019, 55811150, 264448, 264369, 264288, 264686, 264768, 56181562, 21906768, 21906769, 55811957, 35695917, 285022, 60170816, 33857023, 65274620, 18108365, 263987, 33857109, 33857349, 35695763, 264628, 18108376, 55811576, 65274791, 35695855, 56182323, 63373044, 60432113, 264563, 264564, 264567, 264569
368	79607265 (735, 736)	Novel Protein sim. GBank gi 3913029 sp P94987 ALR_MYCSM - ALANINE RACEMASE	UNCLASSIFIED	264508, 264604, 264605, 264638
369	95292917 (737, 738)			
370	88090966 (739, 740)	Novel Protein sim. GBank gi 3249559 (AF018261) - EH domain binding protein Epsin [Rattus norvegicus]		264905, 264592, 264605, 264766, 264691
371	95292599 (741, 742)	Novel Protein sim. GBank gi 2995299 emb CAA18328  - (AL022268) putative RNA delta(2)-isopentenylpyrophosphate transferase [Streptomyces coelicolor]	Contains protein domain (PF01715) - transferase IPP transferase	264905, 264906, 264510, 264600, 264601, 264602, 264603, 265018, 264604, 264605, 265021, 264692, 264636, 264564
372	80021107 (743, 744)	Novel Protein sim. GBank gi 2508393 sp P31576 FIXX_ECOLI - FERREDOXIN LIKE PROTEIN		264584
373	78663766 (745, 746)			
374	78847568 (747, 748)	Novel Protein sim. GBank gi 3341640 emb CAA13164  - (AJ231122) 261f [Vibrio cholerae]	UNCLASSIFIED	264909
375	91230181 (749, 750)	Novel Protein sim. GBank gi 5456934 gb AAD43716.1  - (AF152322) protocadherin gamma A2 [Homo sapiens]	UNCLASSIFIED	264905, 264906
376	80505214 (751, 752)	Novel Protein sim. GBank gi 1805408 dbj BAA08970  - (D50453) homologues to nitrite hydratase region 3'-hypothetical protein P47K of P. chlororaphis [Bacillus subtilis]	cadherin	63274572, 284259, 29331826, 56182435, 60433356, 60433438, 284757, 55812038, 264758, 55811957, 264690, 33857023, 264769
377	10339083 (753, 754)		UNCLASSIFIED	264906

376	80056153 (755, 756)	Novel Protein sim. GBank gi1076013 prj A-9930 - carB (fragment)	Contains protein domain (PF00289) - synthase	UNCLASSIFIED	265008, 264555
378	800503437 (757, 758)	Novel Protein sim. GBank gi216556 dbj BA002174 - (D12651) glucose dehydrogenase [Escherichia coli]	Contains protein domain (PF01011) - dehydrogenase	UNCLASSIFIED	264769
380	80060937 (759, 760)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264604
381	11766027 (761, 762)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264684
382	80054377 (763, 764)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
383	83256025 (765, 766)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
384	95314255 (767, 768)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
385	10237678 (769, 770)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
386	79633434 (771, 772)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
387	17960637 (773, 774)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
388	87741376 (775, 776)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
389	79316971 (777, 778)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
390	80079949 (779, 780)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
391	7657302 (781, 782)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
392	78766056 (783, 784)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
393	33206031 (785, 786)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
394	10104463 (787, 788)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
395	80229010 (789, 790)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
396	20436224 (791, 792)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
397	80417014 (793, 794)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592
398	91230517 (795, 796)	Novel Protein sim. GBank gi13327136 dbj BAA31636 - (AB014561) KIAA0661 protein [Homo sapiens]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenase	UNCLASSIFIED	264592

399	60055278 (797, 798)	Novel Protein sim. GBank gi 3356091 dbj BAA319951 - (AB015974) glycerol kinase [Pseudomonas tolaasii]	Contains protein domain (PF00370) - kinases FGGY family of carbohydrate kinases	kinase	264592, 264595
400	94117480 (789, 800)	Novel Protein sim. GBank gi 728835 sp P39182 ALUS_HUMAN - !!! ALU SUBFAMILY SC WARNING ENTRY !!!	Contains protein domain (PF00560) - Leucine Rich Repeat	cadherin	18108394, 56182575, 22278995, 22278997, 22278998, 264259, 28331824, 265006, 265007, 265008, 60432228, 33657402, 21906754, 265010, 265017, 265018, 265019, 18108351, 18108357, 21908765, 265021, 265022, 264691, 264692, 33657023, 18108370, 65274791, 264634, 264636, 60170394, 56182323, 264594
401	11397481 (801, 802)	Novel Protein sim. GBank gi 4928292 gb AAD33527.1 AF132111 - (AF132117) FhuA [Staphylococcus aureus]		transport	
402	95420284 (803, 804)	Novel Protein sim. GBank gi 5689487 dbj BAA83027.1 - (AB028998) KIAA1075 protein [Homo sapiens]	Contains protein domain (PF00017) - Src homology domain 2	phosphatase	65274572, 56182575, 35698288, 22278998, 22278998, 264093, 264259, 29331822, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 264906, 264907, 264908, 265008, 264511, 265007, 265008, 264910, 264591, 33637402, 60433356, 80433438, 264596, 21906754, 52644296, 265010, 265011, 87188558, 265017, 265018, 265019, 284681, 18108351, 284882, 284448, 264288, 264684, 264766, 264767, 264688, 21906765, 21906766, 21908767, 21908768, 21906769, 55811957, 265020, 265021, 265022, 264690, 264693, 65274620, 35695763, 264628, 18108370, 264629, 18108378, 35696423, 55811576, 264635, 264636, 264557, 264639, 18108385, 22279002, 264563, 264564, 264595, 264568, 264768, 264632, 264639, 264563, 264682, 265009, 264682
403	80439913 (805, 806)			UNCLASSIFIED	
404	11608865 (807, 808)			polymerase	
405	79471280 (809, 810)	Novel Protein sim. GBank gi 2661649 emb CAA157551 - (AL009198) dnaE2 [Mycobacterium tuberculosis]			
406	79634172 (811, 812)		Contains protein domain (PF00159) - Pancreatic hormone peptides	UNCLASSIFIED	18108357, 264693
407	80478229 (813, 814)			UNCLASSIFIED	264769
408	80079556 (815, 816)			UNCLASSIFIED	264600
409	5640527 (817, 818)	Novel Protein sim. GBank gi 3047117 (AF058919) - similar to ATP-dependent RNA helicases [Arabidopsis thaliana]		helicase	264259

410	95357495 (819, 820)	Novel Protein sim. GBank gij475016[dbj BAA05184] - (D29801) Unknown [Mus musculus]	UNCLASSIFIED	264488, 52646365, 52848842, 56181886, 35596286, 52645080, 28331822, 28331824, 56182181, 29331825, 60424289, 35596052, 33556970, 264508, 264509, 264905, 264906, 264907, 264908, 52644045, 264909, 264510, 285007, 264512, 265008, 264910, 33657402, 264756, 52646317, 55811386, 265010, 265011, 265017, 264604, 265018, 55811150, 264762, 264764, 264766, 264687, 264768, 264769, 52644229, 21806766, 265020, 285021, 264534, 52644150, 264682, 33557023, 65274620, 33857109, 33857182, 27486281, 35595763, 264828, 264829, 60431528, 18108378, 263978, 35598423, 35598555, 264632, 264634, 264635, 264637, 264638, 264598, 264639, 56182323, 264599, 60432113, 22279002, 284563, 264565, 264486
411	80501670 (821, 822)	Novel Protein sim. GBank gij3251784[emb CAB08957] - (Z95558) hpx [Mycobacterium tuberculosis]	UNCLASSIFIED	284789
412	80241662 (823, 824)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	eph	284907, 264810, 263873, 22279002, 264605
413	11076446 (825, 826)	Novel Protein sim. GBank gij4668350[gb AAD31273.1 AF13202 - (AF132025) rhophilin [Drosophila melanogaster]	dehydrogenase	18108374, 264760, 264769, 264602, 284638, 264603, 264908, 264605
414	82050554 (827, 828)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	UNCLASSIFIED	264908, 87168518
415	84453144 (829, 830)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	kinase	264488, 264600, 264602, 264764, 264636, 264605
416	80402775 (831, 832)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	Contains protein domain (PF00145) - C-5 cytosine-specific DNA methylase	264605
417	20153787 (833, 834)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	UNCLASSIFIED	284680, 264603
418	84125841 (835, 836)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	collagen	264908, 264810, 264764, 264639
419	95314273 (837, 838)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	phosphatase	264769
420	37036349 (839, 840)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	Contains protein domain (PF00300) - Phosphoglycerate mutase family	264905, 264600, 264601, 264603, 264604, 264760, 264769
421	95292942 (841, 842)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	Contains protein domain (PF00072) - Response regulator receiver domain	22278996, 264682, 18108376, 18108387
422	79471293 (843, 844)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	Contains protein domain (PF00118) - TCP-1/cpn80 chaperonin family	
423	79604948 (845, 846)	Novel Protein sim. GBank gij129038[sp P20707 ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	UNCLASSIFIED	284509

424	78986557 (847, 848)	Novel Protein sim. GBank gi 4026814 refNP_004977.1 pKTN1 - kinesin 1 (kinesin receptor)		struct	265019
425	80431450 (849, 850)	Novel Protein sim. GBank gi 1703701 bbsl178462 - KRP5=kinesin-related protein [rats, testes, Peptide Partial, 167 aa]	Contains protein domain (PF00225) - Kinesin motor domain	struct	264908, 265007, 55811386, 264768, 55810764
426	80064522 (851, 852)	Novel Protein sim. GBank gi 231829 sp P29929 COBN_PSEDE - COBN PROTEIN		UNCLASSIFIED	264605, 264559
427	80057232 (853, 854)	Novel Protein sim. GBank gi 231829 sp P29929 COBN_PSEDE - COBN PROTEIN		UNCLASSIFIED	264603, 264636
428	78487798 (855, 856)	Novel Protein sim. GBank gi 81286 pir S22897 - extensin - Volvox carterii (fragment)		UNCLASSIFIED	264683
429	80091252 (857, 858)	Novel Protein sim. GBank gi 1806154 emb CA806451  - (Z8+385) hypothetical protein Rv0688 [Mycobacterium tuberculosis]		UNCLASSIFIED	35696423, 35695763, 35695855, 265017, 264564, 264762
430	80504192 (859, 860)	Novel Protein sim. GBank gi 1806154 emb CA806451  - (Z8+385) hypothetical protein Rv0688 [Mycobacterium tuberculosis]		reductase	264508, 264905, 264509, 264908, 264909, 265008, 264600, 264687, 264769, 264689, 264636, 264638, 18108385, 264486
431	20624249 (861, 862)	Novel Protein sim. GBank gi 3123552 emb CAA18609  - (ALD22578) dJ393P12.2 (hypothetical Prolin-rich protein KIAA0268 LIKE) [Homo sapiens]		UNCLASSIFIED	264566
432	16525372 (863, 864)	Novel Protein sim. GBank gi 3123552 emb CAA18609  - (ALD22578) dJ393P12.2 (hypothetical Prolin-rich protein KIAA0268 LIKE) [Homo sapiens]		UNCLASSIFIED	265020
433	81494303 (865, 866)	Novel Protein sim. GBank gi 2495272 sp Q99626 CDX2_HUMAN - HOMEBOX PROTEIN CDX-2 (CAUDAL-TYPE HOMEBOX PROTEIN 2) [CDX-3]	Contains protein domain (PF00169) - PH domain	UNCLASSIFIED	264807, 264908, 264909, 264910, 264592, 264595, 264758, 264604, 264760, 264762, 264763, 264636, 264637, 22279002
434	84326323 (867, 868)	Novel Protein sim. GBank gi 2495272 sp Q99626 CDX2_HUMAN - HOMEBOX PROTEIN CDX-2 (CAUDAL-TYPE HOMEBOX PROTEIN 2) [CDX-3]		UNCLASSIFIED	55812038, 56182181, 56181562, 29331828, 35696052, 55810764, 55811578, 65274791, 35695855, 60432113, 55811150, 264636, 264766
435	80502738 (869, 870)	Novel Protein sim. GBank gi 114105 sp P08532 ARAH_ECOLI - L-ARABINOSE TRANSPORT SYSTEM PERMEASE PROTEIN ARAH		transport	264595, 264769
436	41085953 (871, 872)	Novel Protein sim. GBank gi 114105 sp P08532 ARAH_ECOLI - L-ARABINOSE TRANSPORT SYSTEM PERMEASE PROTEIN ARAH		UNCLASSIFIED	265020, 22279002
437	11399291 (873, 874)	Novel Protein sim. GBank gi 114105 sp P08532 ARAH_ECOLI - L-ARABINOSE TRANSPORT SYSTEM PERMEASE PROTEIN ARAH		UNCLASSIFIED	264593
438	11773635 (875, 876)	Novel Protein sim. GBank gi 114105 sp P08532 ARAH_ECOLI - L-ARABINOSE TRANSPORT SYSTEM PERMEASE PROTEIN ARAH		UNCLASSIFIED	264686
439	80019495 (877, 878)	Novel Protein sim. GBank gi 3242702 (AC003040) - hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	264905, 264600, 264602, 264604
440	79841062 (879, 880)	Novel Protein sim. GBank gi 2291232 gb AA85351.1  - (AF016427) Contains similarity to Pfam domain: PF00004 (AAA), Score=268.1, E-value=3.7e-77, N=1 [Caenorhabditis elegans]	Contains protein domain (PF00004) - ATPases associated with various cellular activities (AAA)	ATPase-associated	35696052, 264905, 264808, 264909, 265011, 35696423
441	20396935 (881, 882)	Novel Protein sim. GBank gi 5639946 gb AA045904.1 AF16132 - (AF161328) histidine kinase CstS [Corynebacterium diptheriae]			264605
442	85281058 (883, 884)	Novel Protein sim. GBank gi 1184780 (U46088) - von Ebner minor salivary gland protein [Mus musculus]		UNCLASSIFIED	29331830, 264909
443	82456427 (885, 886)	Novel Protein sim. GBank gi 5639946 gb AA85351.1  - (AL108732) putative ATP-binding RNA helicase [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	35696052, 264508, 264808, 264512, 264604, 264782, 264769, 264689, 264636
444	11395897 (887, 888)	Novel Protein sim. GBank gi 1783249 dbj BA11726  - (D83026) homologous to citrate-sodium symport (citrate transporters); hypothetical [Bacillus subtilis]		UNCLASSIFIED	264591

445	79552709 (889, 890)	Novel Protein sim. GBank gii5531272[emb]CAB50897.1] - (AJ243800) WSC4 homologue [Kluyveromyces fragilis]		UNCLASSIFIED	264593
446	79810937 (891, 892)	Novel Protein sim. GBank gii538413 (L36315) - zinc finger protein [Mus musculus]	Contains protein domain (PF00098) - Zinc finger, C2H2 type	transcript factor	264509
447	80433886 (893, 894)	Novel Protein sim. GBank gii1542914[emb]CAB02185] - (Z80108) fml [Mycobacterium tuberculosis]	Contains protein domain (PF00551) - Formyl transferase	dehydrogenase	264768, 55811576
448	80238110 (895, 896)	Novel Protein sim. GBank gii118794[sp]P10443[DP3A_ECOLI - DNA POLYMERASE III, ALPHA CHAIN		polymerase	264508, 264600, 264603, 264605, 264682, 264769, 18108362, 264634, 18108387
449	20460634 (897, 898)	Novel Protein sim. GBank gii4589506[db]BAA76775.1] - (AB023148) KIAA0931 protein [Homo sapiens]	Contains protein domain (PF00481) - Protein phosphatase 2C	phosphatase	264605, 264559
450	94631210 (899, 900)	Novel Protein sim. GBank gii2493000[sp]Q08450[SCOT_CAEEL - PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE PRECURSOR (3-OXOACID COA-TRANSFERASE)		UNCLASSIFIED	55274572, 22278988, 29331824, 29331826, 264906, 264910, 264592, 52646317, 265017, 21806767, 55811857, 56526486, 22279002
451	21433609 (901, 902)	Novel Protein sim. GBank gii2650614 (AE001104) - conserved hypothetical protein [Archaeoglobus fulgidus]		UNCLASSIFIED	264486
452	10267276 (903, 904)	Novel Protein sim. GBank gii2493000[sp]Q08450[SCOT_CAEEL - PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE PRECURSOR (3-OXOACID COA-TRANSFERASE)		UNCLASSIFIED	264692
453	52660098 (905, 906)	Novel Protein sim. GBank gii4468899[emb]CAB38153.1] - (AL035591) putative integral membrane export protein [Streptomyces coelicolor]		UNCLASSIFIED	264907, 264600
454	39523922 (907, 908)	Novel Protein sim. GBank gii3411053 (AF034863) - synaptic scaffolding molecule [Rattus norvegicus]		transferase	264603
455	13086682 (909, 910)	Novel Protein sim. GBank gii4506075[ref]NP_002733.1[ppRKC - protein kinase C, mu 1]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF)	UNCLASSIFIED	264687
456	79563081 (911, 912)	Novel Protein sim. GBank gii113158[sp]P25516[ACO1_ECOLI - ACONITASE HYDRATASE 1 (CITRATE HYDRO-LYASE 1) (ACONITASE 1)]	Contains protein domain (PF00130) - Phorbol ester/diacylglycerol binding domain (C1 domain)	UNCLASSIFIED	264691
457	79831273 (913, 914)	Novel Protein sim. GBank gii1168574[sp]P42464[ATPB_CORGL - ATP SYNTHASE BETA CHAIN			264905
458	79581227 (915, 916)	Novel Protein sim. GBank gii1168574[sp]P42464[ATPB_CORGL - ATP SYNTHASE BETA CHAIN		kinase	55812038, 265010, 265018, 264681
459	80567359 (917, 918)	Novel Protein sim. GBank gii1168574[sp]P42464[ATPB_CORGL - ATP SYNTHASE BETA CHAIN		kinase	22278997, 264259, 26331826, 265018, 264448, 264389, 21906765, 35596423
460	79245890 (919, 920)	Novel Protein sim. GBank gii1168574[sp]P42464[ATPB_CORGL - ATP SYNTHASE BETA CHAIN		UNCLASSIFIED	264908
461	95287618 (921, 922)	Novel Protein sim. GBank gii1168574[sp]P42464[ATPB_CORGL - ATP SYNTHASE BETA CHAIN		synthase	264602, 264605, 264768, 264769, 265021, 33657023, 264559

462	79605589 (923, 924)	Novel Protein sim. GBank gi1346891 sp P45597 PTF1_XANCP - MULTIPHOSPHORYL TRANSFER PROTEIN (MTP) (CONTAINS: PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I); PHOSPHOCARRIER PROTEIN HPR (PROTEIN H); PTS SYSTEM, FRUCTOSE-SPECIFIC IIA COMPONENT ... (X83413) U88 (Human herpesvirus 6)	Contains protein domain (PF00331) - PEP-utilizing enzymes	UNCLASSIFIED	264807
463	79796417 (925, 926)	Novel Protein sim. GBank gi1854065 emb CAA58337  - (X83413) U88 (Human herpesvirus 6)		UNCLASSIFIED	264805, 264806, 264808, 264809, 264810, 264591, 264595, 265011, 264632, 264635, 264636, 264637, 264638, 264639
464	82340151 (927, 928)	Novel Protein sim. GBank gi15689776 emb CAB32137.1  - (AJ242832) calpain [Homo sapiens]	Contains protein domain (PF00648) - Calpain family cysteine protease	UNCLASSIFIED	264634
465	83005730 (929, 930)	Novel Protein sim. GBank gi1806175 emb CAB06470  - (Z84395) rpsC [Mycobacterium tuberculosis]	Contains protein domain (PF00417) - Ribosomal protein S3, N-terminal domain.	UNCLASSIFIED	265017, 21906764, 265020
466	20460645 (931, 932)	Novel Protein sim. GBank gi1806175 emb CAB06470  - (Z84395) rpsC [Mycobacterium tuberculosis]		UNCLASSIFIED	264805, 264559
467	80405035 (933, 934)	Novel Protein sim. GBank gi1548705 sp P36949 R8SB_BACSU - D-RIBOSE-BINDING PROTEIN PRECURSOR		UNCLASSIFIED	264764
468	52562208 (935, 936)	Novel Protein sim. GBank gi12114024 emb CAB08957  - (Z95558) grcC1 [Mycobacterium tuberculosis]		UNCLASSIFIED	264692
469	19520527 (937, 938)	Novel Protein sim. GBank gi12909459 emb CAA17347  - (AL021929) cobQ [Mycobacterium tuberculosis]		UNCLASSIFIED	264488
470	80502756 (939, 940)	Novel Protein sim. GBank gi114921 sp P17447 BETT_ECOLI - HIGH-AFFINITY CHOLINE TRANSPORT PROTEIN	synthase		264602, 264769
471	17937351 (941, 942)	Novel Protein sim. GBank gi114921 sp P17447 BETT_ECOLI - HIGH-AFFINITY CHOLINE TRANSPORT PROTEIN	transport		265019
472	80047458 (943, 944)	Novel Protein sim. GBank gi1862343 (L10908) - Gcap1 gene product [Mus musculus]		UNCLASSIFIED	264596, 264885, 264557
473	20558793 (945, 946)	Novel Protein sim. GBank gi1862343 (L10908) - Gcap1 gene product [Mus musculus]		UNCLASSIFIED	264369
474	80593365 (947, 948)	Novel Protein sim. GBank gi1862343 (L10908) - Gcap1 gene product [Mus musculus]		UNCLASSIFIED	22278997, 284692, 284288
475	82454665 (949, 950)	Novel Protein sim. GBank gi1862343 (L10908) - Gcap1 gene product [Mus musculus]		UNCLASSIFIED	264907, 264908, 264511, 265008, 264782, 264448, 264636, 264638
476	94143857 (951, 952)	Novel Protein sim. GBank gi15433858 ref NP_008329.1 pGAC1 - glioma amplified on chromosome 1 protein (leucine-rich)	Contains protein domain (PF00560) - Leucine Rich Repeat		65274572, 60432048, 284298, 264508, 52644045, 55812038, 284758, 265011, 284288, 264686, 52644228, 65274791, 284838, 284586
477	79175833 (953, 954)	Novel Protein sim. GBank gi15433858 ref NP_008329.1 pGAC1 - glioma amplified on chromosome 1 protein (leucine-rich)		UNCLASSIFIED	264638
478	78633483 (955, 956)	Novel Protein sim. GBank gi15433858 ref NP_008329.1 pGAC1 - glioma amplified on chromosome 1 protein (leucine-rich)		UNCLASSIFIED	264690, 264693
479	80169746 (957, 958)	Novel Protein sim. GBank gi15433858 ref NP_008329.1 pGAC1 - glioma amplified on chromosome 1 protein (leucine-rich)	collagen		264686, 35695855, 265008, 264631, 264810, 264632, 264638, 265018, 264359, 264909
480	79390729 (959, 960)	Novel Protein sim. GBank gi1127551 (U18939) - orf2 [Batrachococcus baikalensis]	mapolymerase		264369
481	79624578 (961, 962)	Novel Protein sim. GBank gi14053042 (AF068085) - GP900; mucin-like glycoprotein [Cryptosporidium parvum]		UNCLASSIFIED	264693
482	83050811 (963, 964)	Novel Protein sim. GBank gi14053042 (AF068085) - GP900; mucin-like glycoprotein [Cryptosporidium parvum]		UNCLASSIFIED	264908, 264886, 284788, 284693, 55811576, 56182323, 18108385

483	20293306 (965, 966)	Novel Protein sim. GBank gi 2104303 emb CAB08632  (Z95387) hypothetical protein RV2610c [Mycobacterium tuberculosis]	Contains protein domain (PF00534) - Glycosyl transferase group 1	264600
484	11618046 (987, 988)	Novel Protein sim. GBank gi 3450883 (AF083334) - fibroin [Antheraea pernyi]	UNCLASSIFIED	264594
485	80191234 (969, 970)	Novel Protein sim. GBank gi 5042272 emb CAB44526.1  - (AL078618) nuoF, NADH dehydrogenase subunit [Streptomyces coelicolor]	UNCLASSIFIED	264604
486	80059042 (971, 972)	Novel Protein sim. GBank gi 5724778 p AAC55522.2  - (AF012273) rho-type GTPase-activating protein rhoGAP-X-1 [Mus musculus]	Contains protein domain (PF00620) - RhoGAP domain	264589, 21908785, 22270000, 22279002
487	11813339 (973, 974)	Novel Protein sim. GBank gi 3852223 p BAA34471.1  - (AB018294) KIAA0751 protein [Homo sapiens]	kinase	264638
488	91222383 (975, 976)	Novel Protein sim. GBank gi 620971 p J25484 - hydroxyproline-rich glycoprotein - tomato (fragment)	collagen	264688, 66714117, 264768, 18108385, 55811576, 265006, 265008, 265009, 265019, 22279002, 264259, 18108370, 264907, 264764, 56182323, 264288, 264693
489	10867710 (977, 978)	Novel Protein sim. GBank gi 289420 emb CAA17072  - (AL021840) hypothetical protein RV3258c [Mycobacterium tuberculosis]	UNCLASSIFIED	264639
490	95361124 (979, 980)	Novel Protein sim. GBank gi 5262605 p BAA45743.1  - (AL080150) hypothetical protein [Homo sapiens]	UNCLASSIFIED	22278996, 29331822, 29331828, 284107, 264909, 264110, 265009, 264592, 264593, 60433396, 264288, 264693, 263974, 263976, 20281071, 60432113
491	80496412 (981, 982)	Novel Protein sim. GBank gi 289420 emb CAA17072  - (AL021840) hypothetical protein RV3258c [Mycobacterium tuberculosis]	UNCLASSIFIED	264769
492	87421264 (983, 984)	Novel Protein sim. GBank gi 5262605 p BAA45743.1  - (AL080150) hypothetical protein [Homo sapiens]	UNCLASSIFIED	264600
493	11692842 (985, 986)	Novel Protein sim. GBank gi 5262605 p BAA45743.1  - (AL080150) hypothetical protein [Homo sapiens]	UNCLASSIFIED	264638
494	87726604 (987, 988)	Novel Protein sim. GBank gi 2791517 emb CAA16054  - (AL021246) hypothetical protein RV2477c [Mycobacterium tuberculosis]	Contains protein domain (PF00005) - ABC transporter	264489, 35696286, 60432289, 29331828, 35696052, 264509, 264805, 264806, 264907, 264808, 264909, 264510, 264511, 265009, 264910, 33657402, 264762, 264764, 264768, 264769, 264688, 21906765, 21906769, 35695917, 285020, 264693, 33657109, 264629, 35696423, 35696455, 264634, 264638
495	80028599 (989, 990)	Novel Protein sim. GBank gi 2791517 emb CAA16054  - (AL021246) hypothetical protein RV2477c [Mycobacterium tuberculosis]	transport	264602, 264682, 264638
496	78985624 (991, 992)	Repressor (Amino-Terminal Domain) (R1-69)	Contains protein domain (PF01381) - Helix-turn-helix	264601, 265021
497	78946661 (993, 994)	Novel Protein sim. GBank gi 128736 p P2825 p DXH_ECOLI - PYRIDOXAMINE 5'-PHOSPHATE OXIDASE (PNP/PMP OXIDASE)	oxidase	265006



498	88095488 (995, 996)	Novel Protein sim. GBank gi 1145789 (U41662) - neurotigin 2 [Rattus norvegicus]	Contains protein domain (PF00135) - Carboxylesterases	esterase	264259, 29331826, 35696032, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265009, 264910, 264591, 33657402, 264758, 265010, 265011, 264600, 264601, 264605, 264683, 264784, 264766, 264767, 264768, 264887, 264769, 21905767, 33657023, 264693, 264628, 264629, 35696423, 264630, 264632, 264634, 264635, 264637, 264638, 264658, 264639, 18108385, 264563, 264564, 264565, 264566, 264567
499	20438222 (897, 898)	Novel Protein sim. GBank gi 97480 pir S19739 - integral membrane protein - Rhodobacter capsulatus		UNCLASSIFIED	264605
500	11076810 (999, 1000)				264605
501	13418034 (1001, 1002)	Novel Protein sim. GBank gi 5708250 emb CAB32363.1  - (AL108747) putative integral membrane protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	264688
502	80021176 (1003, 1004)	Novel Protein sim. GBank gi 468678 emb CAB38132.1  - (AL035591) glucose-6-phosphate isomerase [Streptomyces coelicolor]	Contains protein domain (PF00342) - Phosphoglucose isomerase	isomerase	22278996, 265011, 264602, 264605, 264635
503	20264483 (1005, 1006)			UNCLASSIFIED	264564
504	10887321 (1007, 1008)			UNCLASSIFIED	264687
505	95003068 (1008, 1010)			UNCLASSIFIED	264369
506	18454292 (1011, 1012)	Novel Protein sim. GBank gi 4033509 sp P02598 CALM_TETPY - CALMODULIN [Novel Protein sim. GBank gi 2501069 sp Q46127 SYW.CLOLO - TRYPTOPHANYL-TRNA SYNTHETASE (TRYPTOPHAN--TRNA LIGASE) (TRPRS)]	Contains protein domain (PF00036) - EF hand	struct	265010
507	20451588 (1013, 1014)	Novel Protein sim. GBank gi 466088 sp P34618 YO82_CAEEL - HYPOTHETICAL 33.8 KD PROTEIN ZK1236.2 IN CHROMOSOME III		UNCLASSIFIED	264604
508	78841424 (1015, 1016)			UNCLASSIFIED	264908
509	11776386 (1017, 1018)				264638
510	83373463 (1019, 1020)			UNCLASSIFIED	264687, 264639
511	18529578 (1021, 1022)				265007
512	20389484 (1023, 1024)	Novel Protein sim. GBank gi 2497419 sp P55635 Y4RB_RHISN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB		UNCLASSIFIED	264565
513	79457404 (1025, 1026)	Novel Protein sim. GBank gi 1276897 (U41809) - cyclin J [Drosophila melanogaster]	Contains protein domain (PF00134) - cyclin	cyclin	264683, 264689, 35696423, 264639
514	79813605 (1027, 1028)	Novel Protein sim. GBank gi 1184780 (U46068) - von Ebner minor salivary gland protein [Mus musculus]		UNCLASSIFIED	29331830, 264909
515	79462581 (1029, 1030)				22278998, 264690
516	9862020 (1031, 1032)	Novel Protein sim. GBank gi 2127400 pir S65770 - malicolicosyltrehalose trehalohydrolase - Arthrobacter sp. (strain Q38)		amylase	264910

517	95292894 (1033, 1034)	Novel Protein sim. GBank gij2983605 (AE000725) - rboose 5 phosphate isomerase B [Aquilifex aeolicus]			isomerase	265018, 264605, 284784, 284766, 284687, 264691, 264565, 284487
518	8491831 (1035, 1036)	Novel Protein sim. GBank gij854065 (emb)CAA583371 - (X83413) U88 [Human herpesvirus 6]			UNCLASSIFIED	
519	91677888 (1037, 1038)	Novel Protein sim. GBank gij5689365 (dbj)BAA83073.1 - (AB024075) B120 [Homo sapiens]		Contains protein domain (PF01388) - ARID DNA binding domain	dna_rna_bind	52844507, 22278997, 22278998, 60432049, 264259, 52845080, 29331824, 86714117, 60424269, 29331826, 35696052, 284805, 29331830, 66712502, 284511, 285007, 264591, 60432229, 33657402, 60433438, 21908754, 33109954, 52844296, 87168474, 87168559, 285017, 285018, 284604, 285018, 264681, 284448, 264369, 284288, 264685, 21908765, 21906766, 21908787, 21908789, 265021, 60170615, 33657023, 284692, 52645129, 33657109, 27486262, 27486264, 35695763, 18108370, 284629, 52844332, 56182323, 284639, 83373044, 18108385, 56526488, 60432113
520	78669188 (1039, 1040)					284789
521	11076821 (1041, 1042)	Novel Protein sim. GBank gij1169126 (sp)P46839 (CTPA, MYCLE - CATION-TRANSPORTING P-TYPE ATPASE A		transport		264605
522	80435060 (1043, 1044)	Novel Protein sim. GBank gij1172869 (sp)P44331 (RBSK, HAEIN - RIBOKINASE		kinase		264905, 284768
523	18358013 (1045, 1046)	Novel Protein sim. GBank gij2132243 (pif)S61028 - hypothetical protein YPL236c - yeast (Saccharomyces cerevisiae)		UNCLASSIFIED		284629
524	80261805 (1047, 1048)	Novel Protein sim. GBank gij4033608 (dbj)BAA35136 - (AB012308) B2HC [Anthracis crassispina]		ATPase associated		264082, 264596, 265011
525	79610048 (1049, 1050)					284907
526	38827630 (1051, 1052)	Novel Protein sim. GBank gij4106610 (emb)CAA21365 - (AL031666) ORF42, len=388 aa, similarity to an aminotransferase, in P95957 Sultolobus solfataricus, (401 aa), 33.1% identity in 393 aa overlap, FastA scores: opt-468, E(-): 8.5e-24, in O64602 R. norvegicus, (425 aa), 28.6% ident...		UNCLASSIFIED		264758
527	80504729 (1053, 1054)					264768
528	65484134 (1055, 1056)					56182575, 265017, 265018
529	17936810 (1057, 1058)	Novel Protein sim. GBank gij731088 (sp)P24215 (UXUA, ECOLI - MANNONATE DEHYDRATASE (D-MANNONATE HYDROLASE)		hydrolase		265019
530	10887336 (1059, 1060)	Novel Protein sim. GBank gij42144 (emb)CAA23200 - (X00513) NusaA protein (nusa) [Escherichia coli]				264687
531	80226575 (1061, 1062)					
532	90833444 (1063, 1064)	Novel Protein sim. GBank gij5262840 (emb)CAB45758.1 - (AL080170) hypothetical protein [Homo sapiens]		Contains protein domain (PF00622) - SPRY domain	UNCLASSIFIED	284555, 264556, 284557, 284558, 18108385, 284488, 284490, 264258, 264592, 264760, 265021, 264690, 263976, 264558

533	87761531 (1065, 1066)	Novel Protein sim. GBank gi 4883838 gb AAD31593.1 AF112299  integral inner nuclear membrane protein MAN1 [Homo sapiens]			264907, 264909, 264768, 35695917, 264630, 264455
534	82368264 (1067, 1068)	Novel Protein sim. GBank gi 2895352 emb CAA04606.1  - (AJ001206) pep1 [Streptomyces coelicolor]	UNCLASSIFIED		264905, 265011, 264601, 264602, 264605, 264762, 264768, 265020, 264693, 264638
535	79641850 (1069, 1070)	Novel Protein sim. GBank gi 3878638 emb CAA08853  - (Z49126) similar to cAMP-dependent protein kinase: cDNA EST EMBL: T00719 comes from this gene: cDNA EST yk465d8.3 comes from this gene: cDNA EST yk465d8.5 comes from this gene: cDNA EST yk492d4.3 comes from this gene: cDNA EST y...	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	- ATPase-associated	264906
536	79907207 (1071, 1072)	Novel Protein sim. GBank gi 2495828 sp P55757 VOHL_SERMA - HYPOTHETICAL 10.1 KD PROTEIN IN BIOA 5 REGION	reductase		18108376, 264905, 264906, 264907, 264909
537	94147448 (1073, 1074)	Novel Protein sim. GBank			265008, 264605, 65274791
538	87821863 (1075, 1076)	gi 134920 sp P21997 SSGP_VOLCA - SULFATED SURFACE GLYCOPROTEIN 185 (SSG 185)	collagen	Contains protein domain (PF00585) - PDZ domain (Also known as DHR or GLGF).	29331822, 29331824, 29331825, 29331826, 29331827, 264908, 52644045, 33657402, 265017, 264762, 264883, 264288, 264685, 21906765, 35695763, 264558, 60170394, 264559, 22279002
539	28396268 (1077, 1078)	Novel Protein sim. GBank gi 2498433 sp Q1234.1 HAT1_YEAST - HISTONE ACETYLTRANSFERASE	histone		264602, 265019
540	78637077 (1079, 1080)	Novel Protein sim. GBank gi 3882241 db BAA34480.1  - (AB018303) KIAA0760 protein [Homo sapiens]	transcriptfactor	Contains protein domain (PF00096) - Zinc finger, C2H2 type	264693
541	87762268 (1081, 1082)	Novel Protein sim. GBank gi 5042272 emb CAB44526.1  - (AL078818) nuoF, NADH dehydrogenase subunit [Streptomyces coelicolor]	dehydrogenase		18108394, 22278997, 22278998, 264259, 264112, 265009, 33657402, 55812038, 52646317, 265017, 21906765, 264693, 55911576, 264635, 56526486, 264568
542	95295836 (1083, 1084)	Novel Protein sim. GBank gi 5042272 emb CAB44526.1  - (AL078818) nuoF, NADH dehydrogenase subunit [Streptomyces coelicolor]			264910, 265018, 264689, 264638, 264488
543	79796290 (1085, 1086)	Novel Protein sim. GBank gi 2791398 emb CAA15884  - (AL021184) hypothetical protein Rv1464 [Mycobacterium tuberculosis]	UNCLASSIFIED		264602, 264908
544	20437191 (1087, 1088)	Novel Protein sim. GBank	UNCLASSIFIED		264605
545	80434504 (1089, 1090)	Novel Protein sim. GBank			264768, 264634, 264907, 264592, 264809
546	80249018 (1091, 1092)	gi 488721 gb AAD32237.1 AF14744 - (AF147449) penicillin binding protein 1B [Pseudomonas aeruginosa]			264600, 264602, 21908785
547	11077563 (1093, 1094)	Novel Protein sim. GBank gi 1350855 sp P19176 RPOC_PSEPU - DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (TRANSCRIPTASE BETA' CHAIN) (RNA POLYMERASE BETA' SUBUNIT) related protein: KRP: Costal2 [Drosophila melanogaster]	mapolymerase		264604
548	82114936 (1095, 1096)	Novel Protein sim. GBank gi 2330021 AFD19250  - kinesin- related protein: KRP: Costal2 [Drosophila melanogaster]	UNCLASSIFIED		264488, 264905, 284910, 284760, 264693, 264639, 264583, 264584

549	95421904 (1097, 1098)	Novel Protein sim. GBank gij4337460jgbjAAD181331 - (AF056195) neuroblastoma-amplified protein (Homo sapiens)		UNCLASSIFIED	264488, 65274572, 18108398, 22278895, 22278998, 22278997, 22278998, 22278999, 264299, 29331824, 66714117, 29331825, 28331826, 35686052, 265007, 265008, 264810, 264592, 33657402, 33109954, 265017, 265018, 265019, 18108351, 264448, 264784, 264369, 264286, 264766, 264686, 264688, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 264691, 33657023, 284692, 284693, 65274620, 52545128, 33657108, 27486261, 27486262, 27486264, 33657348, 55811576, 18108387, 60432113, 22279002, 264688
550	10866516 (1099, 1100)	Novel Protein sim. GBank		UNCLASSIFIED	264908, 264809, 264768
551	80439990 (1101, 1102)	gij3122893spjP94985(SYFB_MYCTU - PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (PHENYLALANINE-TRNA LIGASE BETA CHAIN) (PHERS)		UNCLASSIFIED	264689, 264638, 284583, 55811957, 264628
552	94672870 (1103, 1104)	Novel Protein sim. GBank		UNCLASSIFIED	264908
553	80106002 (1105, 1106)	Novel Protein sim. GBank protein [Drosophila melanogaster]	Contains protein domain (PF00008) - EGF-like domain	glycoprotein	
554	79618378 (1107, 1108)	Novel Protein sim. GBank gij5019771jgbjAAD37857.1IAF13326 - (AF133263) histidine protein kinase-response regulator hybrid protein CvgSY [Pseudomonas syringae pv. syringae]		kinase	
555	78996347 (1109, 1110)	Novel Protein sim. GBank gij131515spjP02508IPTGA_SALTY - PTS SYSTEM, GLUCOSE-SPECIFIC IIA COMPONENT (EIIA-GLC) (GLUCOSE-PERMEASE IIA COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, A COMPONENT) (EIIA-GLC)	Contains protein domain (PF00358) - phosphoenolpyruvate-dependent sugar phosphotransferase system, EIIA 1	transport	264762
556	20457127 (1111, 1112)	Novel Protein sim. GBank gij3914014spjP96380MFD_MYCTU - TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF)		transcript factor	284508, 284605, 284559
557	19523405 (1113, 1114)	Novel Protein sim. GBank gij5042273jembjCAB44527.1 - (AL078818) nuoE, NADH dehydrogenase subunit [Streptomyces coelicolor]		dehydrogenase	264488
558	20724428 (1115, 1116)	Novel Protein sim. GBank gij1170933spjP45331METE_HAEIN - 5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE-HOMOCYSTEINE METHYLTRANSFERASE (METHIONINE SYNTHASE, VITAMIN-B12 INDEPENDENT ISOZYME) (COBALAMIN-INDEPENDENT METHIONINE SYNTHASE)		UNCLASSIFIED	264602
559	80084353 (1117, 1118)	Novel Protein sim. GBank gij4980567jgbjAAD35173.1IAE00169 - (AE001694) iron(III) ABC transporter, permease protein [Thermotoga maritima]		UNCLASSIFIED	284634

560	80066533 (1118, 1120)	Novel Protein sim. GBank gi 2492595 sp Q53193 Y4TR_RHISN - PROBABLE PEPTIDE ABC TRANSPORTER ATP-BINDING PROTEIN Y4TR	Contains protein domain (PF00005) - ABC transporter	transport	18108396, 264806, 264802, 264604, 18108374
561	20293187 (1121, 1122)			UNCLASSIFIED	264600
562	11698161 (1123, 1124)			UNCLASSIFIED	264689
563	79781420 (1125, 1126)	Novel Protein sim. GBank gi 4104925 (AF042276) - poly(hydroxycarboxylate) granule associated protein GA2 [Pseudomonas putida]		UNCLASSIFIED	264910, 264691
564	56716390 (1127, 1128)	Novel Protein sim. GBank gi 2792310 (AF040570) - unknown [Amycolatopsis mediterranei]		dehydrogenase	264592
565	56465618 (1129, 1130)	Novel Protein sim. GBank gi 3449294 dbj BAA32462  - (AB011532) MEGF8 [Rattus norvegicus]	Contains protein domain (PF00008) - EGF-like domain	synthase	265010
566	94323888 (1131, 1132)	Novel Protein sim. GBank gi 4539569 emb CAB38487.1  - (AL035638) putative helicase [Streptomyces coelicolor]		helicase	264908, 264510, 265008, 264910, 264758, 264600, 264602, 264604, 264605, 264768, 264687, 264689, 35695917, 264693, 65274620, 264488
567	79560955 (1133, 1134)	Novel Protein sim. GBank gi 100506 pir IS17455 - Malate dehydrogenase (oxalacetate-decarboxylating) (NADP+) (EC 1.1.1.40) - Flaveria trinervia (fragment)	Contains protein domain (PF00390) - Malic enzyme	UNCLASSIFIED	264681, 264691, 264593
568	94681793 (1135, 1136)	Novel Protein sim. GBank gi 3915843 sp O31212 RS2_STRCO - 30S RIBOSOMAL PROTEIN S2	Contains protein domain (PF00318) - Ribosomal protein S2	ribosomalprot	264565
569	39508897 (1137, 1138)	Novel Protein sim. GBank gi 115122 sp P21627 BRAD_PSEAE - HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PROTEIN BRAD		UNCLASSIFIED	18108376, 18108387, 264565
570	78375927 (1139, 1140)			transport	264907, 264909
571	79793961 (1141, 1142)				
572	36996838 (1143, 1144)	Novel Protein sim. GBank gi 4539223 emb CAB39881.1  - (AL049497) putative Integral membrane protein [Streptomyces coelicolor]		UNCLASSIFIED	264782
573	20716521 (1145, 1146)			UNCLASSIFIED	265007, 264601
574	13521592 (1147, 1148)				264636
575	13076416 (1149, 1150)	Novel Protein sim. GBank gi 118794 sp P10443 DP3A_ECOLI - DNA POLYMERASE III, ALPHA CHAIN		polymerase	264687
576	20482246 (1151, 1152)	Novel Protein sim. GBank gi 5457625 emb CAB49116.1  - (AJ248283) PAB2227 [Pyrococcus abyssi]			264605
577	56727102 (1153, 1154)	Novel Protein sim. GBank gi 5042274 emb CAB44528.1  - (AL078618) nuoD, NADH dehydrogenase subunit [Streptomyces coelicolor]	Contains protein domain (PF00346) - Respiratory-chain NADH dehydrogenase, 49 Kd subunit	dehydrogenase	35696052, 264636
578	11804477 (1155, 1156)				264638
579	11794723 (1157, 1158)	Novel Protein sim. GBank gi 1723081 sp Q11046 Y089_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.09		transport	264682, 264556

580	80059417 (1159, 1160)				22276999, 35696052, 264555, 264556, 264558
581	79230833 (1161, 1162)			UNCLASSIFIED	265008, 264564
582	80049617 (1163, 1164)		Contains protein domain (PF000047) - Immunoglobulin domain	struct	265021, 264555, 264557
583	79321392 (1165, 1166)			transport	264594
584	78845024 (1167, 1168)				
585	78581454 (1168, 1170)			UNCLASSIFIED	264488, 264906, 264766, 264887, 35698423
586	38277485 (1171, 1172)			UNCLASSIFIED	265018, 264684, 21906769
587	80497359 (1173, 1174)			UNCLASSIFIED	264908, 265007
588	78557239 (1175, 1176)			hydrolase	264600, 264602, 264605, 264769, 264690, 264557
589	79805628 (1177, 1178)			UNCLASSIFIED	265020, 264692
590	78815628 (1179, 1180)			UNCLASSIFIED	22276999, 264807, 264809, 264510, 285008, 265010, 264687, 264769, 35695917, 18108376, 264634, 264638, 264638
591	10313540 (1181, 1182)			UNCLASSIFIED	264906, 264909
592	13889767 (1183, 1184)			mapolymerase	264691
593	82348698 (1185, 1186)			MHC	263972
594	20212392 (1187, 1188)			dehydrogenase	264511, 264762, 264769, 264486
595	10064064 (1189, 1190)			UNCLASSIFIED	264605
596	13085170 (1191, 1192)				264769
597	80259003 (1193, 1194)			UNCLASSIFIED	264636
598	84140216 (1195, 1196)			UNCLASSIFIED	264592
599	20385137 (1197, 1198)			UNCLASSIFIED	264758, 55810764, 264555, 264556, 264637, 83373044
600	10357693 (1199, 1200)			UNCLASSIFIED	264603
601	79810404 (1201, 1202)			UNCLASSIFIED	264908, 264510

602	79250602 (1203, 1204)	Novel Protein sim. GBank gij352286 [gbjAAC34243.1] - (AC004411) putative pro kinase [Arabidopsis thaliana]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	265007
603	11466087 (1205, 1206)			UNCLASSIFIED	264595
604	81675420 (1207, 1208)				264758
605	20436657 (1209, 1210)	Novel Protein sim. GBank gij1175322 [spjP44917]Y883_HAEIN - HYPOTHETICAL PROTEIN H0883		UNCLASSIFIED	264505
606	80334582 (1211, 1212)	Novel Protein sim. GBank gij5020264 [gbjAAD38043.1]AF15136 - (AF151363) Cdc42 GTPase-activating protein [Mus musculus]		UNCLASSIFIED	264764
607	95361506 (1213, 1214)	Novel Protein sim. GBank gij188864 (M74027) - mucin [Homo sapiens]		UNCLASSIFIED	264508, 264906, 85658542, 264682, 264687, 264689, 264534, 18108376, 35696423, 264636, 264555, 264638
608	11810888 (1215, 1216)			UNCLASSIFIED	264882
609	80064775 (1217, 1218)	Novel Protein sim. GBank gij249670 [spjP55552]Y4LL_RHISN - HYPOTHETICAL 91.8 KD PROTEIN Y4LL	Contains protein domain (PF00989) - PAS domain	UNCLASSIFIED	264805
610	79629413 (1219, 1220)				264692
611	87586205 (1221, 1222)				264508, 264905, 264907, 264908, 264909, 264511, 264910, 264758, 264604, 264684, 264766, 264689, 264692, 264628, 264635, 264636, 264637, 264558
612	95287651 (1223, 1224)	Novel Protein sim. GBank gij1877368 [embjCAB07118] - (Z92772) recD [Mycobacterium tuberculosis]	Contains protein domain (PF01443) - Viral (Superfamily 1) RNA helicase	nuclease	264600, 264601, 264604, 264769, 264558, 264565
613	7523475 (1225, 1226)			UNCLASSIFIED	264369
614	78969348 (1227, 1228)	Novel Protein sim. GBank gij5114231 [gbjAAD40238.1]AF13670 - (AF136709) histidine kinase YycG [Staphylococcus aureus]		kinase	18108372, 264563
615	38586986 (1229, 1230)	Novel Protein sim. GBank gij133950 [gbjBAA12741] - (D65230) large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]		synthase	264600, 264602, 264629
616	20465331 (1231, 1232)	Novel Protein sim. GBank gij544367 [spjP35673]GALE_ERWAM - UDP-GLUCOSE 4-EPIMERASE (GALACTOWALDENASE) (UDP-GALACTOSE 4-EPIMERASE)		isomerase	264605
617	81227222 (1233, 1234)	Novel Protein sim. GBank gij2498097 [spjQ60759]TNP3_MOUSE - TUMOR NECROSIS FACTOR, ALPHA-INDUCED PROTEIN 3 (PUTATIVE DNA BINDING PROTEIN A20) (ZINC FINGER PROTEIN A20)	Contains protein domain (PF00641) - Inf Zn-finger in Ran binding protein and others.	Inf	52645156, 21908765, 35696423, 21908786, 21906769, 22278994, 35696286, 22278996, 265020, 265021, 265007, 265008, 264636, 52644150, 33657023, 264692, 264693, 29331822, 29331824, 55812038, 83373044, 56182181, 60424269, 66714117, 29331825, 33657109, 29331826, 33657182, 29331827, 35696052, 29331828, 27486262, 33657349, 56526486, 265018, 265019, 22279002, 264482, 264448, 29331830, 66712502, 264909

618	20632843 (1235, 1236)	Novel Protein sim. GBank gi 5459388 emb CAB50746.1  - (AL096839) putative aminotransferase (Streptomyces coelicolor)		isomerase	264603
619	91227224 (1237, 1238)				56994075, 29331826, 33656970, 265008, 33657402, 33109954, 87168559, 284448, 18108374, 83373044
620	81183143 (1239, 1240)	Novel Protein sim. GBank gi 464335 sp Q05922 DUS2_MOUSE - DUAL SPECIFICITY PROTEIN PHOSPHATASE 2 (DUAL SPECIFICITY PROTEIN PHOSPHATASE PAC-1)		phosphatase	28146498, 264758, 264369, 29148527
621	80239251 (1241, 1242)	Novel Protein sim. GBank gi 2633557 emb CAB13060  - (Z99110) yjdf [Bacillus subtilis]		UNCLASSIFIED	264556, 264558, 264639
622	20456427 (1243, 1244)	Novel Protein sim. GBank gi 185771 gb AA848482  - (U87224) contactin associated protein [Rattus norvegicus]		UNCLASSIFIED	264605
623	10131788 (1245, 1246)	Novel Protein sim. GBank gi 1705703 sp P52225 COMF_PSEFL - CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYCK	Contains protein domain (PF000054) - laminin Laminin G domain	laminin	284908
624	19534127 (1247, 1248)	Novel Protein sim. GBank gi 1705703 sp P52225 COMF_PSEFL - CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYCK		cytochrome	264586
625	13084619 (1249, 1250)	Novel Protein sim. GBank gi 2894252 emb CAA17114.1  - (AL021841) hypothetical protein Rv3342 [Mycobacterium tuberculosis]		UNCLASSIFIED	264688
626	88062603 (1251, 1252)	Novel Protein sim. GBank gi 416592 sp P32323 AGA1_YEAST - A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR		UNCLASSIFIED	28131822, 264905, 264808, 33657023, 33657109, 264558
627	80255457 (1253, 1254)	Novel Protein sim. GBank gi 3098418 AF040944  - P140 [Mus musculus]		UNCLASSIFIED	18108394, 264112, 264583, 265022, 264635
628	80077096 (1255, 1256)	Novel Protein sim. GBank gi 1711543 sp P40526 SSP1_SCHPO - SERINE/THREONINE-PROTEIN KINASE SSP1	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain	kinase	264600
629	79851602 (1257, 1258)	Novel Protein sim. GBank gi 1143204 U34305  - ORF2; Method: conceptual translation supplied by author. [Shigella sonnei]		isomerase	264906, 264907
630	39565156 (1259, 1260)	Novel Protein sim. GBank gi 3236368 AF064748  - S3-12 [Mus musculus]		UNCLASSIFIED	264490
631	20598718 (1261, 1262)	Novel Protein sim. GBank gi 140687 sp P11686 YGG8_ECOLI - HYPOTHETICAL 30.9 KD PROTEIN IN SBM-FBA INTERGENIC REGION (ORF 4) (F288)			263978
632	27843890 (1263, 1264)			UNCLASSIFIED	284908, 264600, 264605, 264769, 264689, 264488
633	80477772 (1265, 1266)			UNCLASSIFIED	264769
634	17938808 (1267, 1268)				265019
635	79574508 (1269, 1270)			UNCLASSIFIED	264689
636	79810981 (1271, 1272)			UNCLASSIFIED	264596, 264762, 264693



637	82455796 (1273, 1274)	Novel Protein sim. GBank gll2326739[emb]CAB10953] - (Z98288) recN [Mycobacterium tuberculosis]		nuclease	264806, 264907, 264510, 264511, 264601, 264602, 264603, 264604, 264605, 18108351, 264762, 264766, 264687, 264769, 264689, 35695917, 264693, 264634, 264638, 264639, 264559, 18108385, 264636
638	14987457 (1275, 1276)	Novel Protein sim. GBank gll4678662[emb]CAB1074.1] - (AL049845) putative large ATP-binding protein [Streptomyces coelicolor]		struct	264112, 263974
639	80204210 (1277, 1278)	Novel Protein sim. GBank gll4589628[db]BAA76836.1] - (AB0233209) KIAA0992 protein [Homo sapiens]		Contains protein domain (PF01466) - rnapolymerase	265009, 265010
640	17929579 (1279, 1280)	Novel Protein sim. GBank gll1432083 (U60981) - homolog to Skp1p, an evolutionarily conserved kinetochore protein in budding yeast [Arabidopsis thaliana]		UNCLASSIFIED	264693
641	79636398 (1281, 1282)			UNCLASSIFIED	264565
642	19898737 (1283, 1284)			UNCLASSIFIED	264906, 264908, 264758, 264288, 264632, 264635, 264639, 264564
643	81516220 (1285, 1286)			UNCLASSIFIED	264684
644	11751367 (1287, 1288)			UNCLASSIFIED	264906, 264762, 264693, 264639, 264559
645	85010907 (1289, 1290)			UNCLASSIFIED	264595, 264566
646	80065083 (1291, 1292)			transcriptfactor	264909, 264591
647	80257085 (1293, 1294)	Novel Protein sim. GBank gll4507613[pe]NP_003738.1]pTNKS - TANKYRASE	Contains protein domain (PF00023) - Ank repeat	UNCLASSIFIED	264600
648	80077428 (1295, 1296)	Novel Protein sim. GBank gll1044983[db]B169846 - protamine [Monodonella turbinata, gonads, Peptide, 106 aa]		UNCLASSIFIED	263978
649	80247447 (1297, 1298)			UNCLASSIFIED	264686
650	11798316 (1299, 1300)			UNCLASSIFIED	264602, 264638
651	11776932 (1301, 1302)	Novel Protein sim. GBank gll1346918[sp]P12283[PURA, ECOLI - ADENYLOSUCCINATE SYNTHETASE (IMP--ASPARTATE LIGASE)]		UNCLASSIFIED	264905, 264907, 264909, 263978, 264637, 22278996, 264510, 264511, 264512, 264593, 21906754, 264603, 264760, 18108376, 264556
652	85516704 (1303, 1304)	Novel Protein sim. GBank gll1722977[sp]Q10638[Y03C_MYCTU - HYPOTHETICAL 82.8 KD PROTEIN CY130.12C]		UNCLASSIFIED	264906, 264595, 264632
653	82124947 (1305, 1306)			transferase	264592
654	85010589 (1307, 1308)	Novel Protein sim. GBank gll130327[sp]P28847[PLSC, ECOLI - 1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (1-AGP ACYLTRANSFERASE) (1-AGPAT) (LYSOPHOSPHATIDIC ACID ACYLTRANSFERASE) (LPAAT)]	Contains protein domain (PF01553) - Acyltransferase	UNCLASSIFIED	264602, 264605, 264766, 264691
655	79320692 (1309, 1310)			UNCLASSIFIED	264557, 264558
656	80418739 (1311, 1312)				
657	20611010 (1313, 1314)				

655	87761815 (1315, 1316)	Novel Protein sim. GBank gi 5689493 dbj BAA03030.1  - (AB029007) KIAA1078 protein (Homo sapiens)		UNCLASSIFIED	22278996, 50432049, 29331822, 29331824, 29331828, 265007, 265009, 33637402, 33637084, 265017, 264448, 21906765, 21906768, 263987, 20281148, 18108370, 18108374, 284482
659	87716663 (1317, 1318)	Novel Protein sim. GBank gi 2137872 pir I48724 - Zinc finger protein PZF - mouse	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	22278999, 60432049, 66714117, 29331827, 265007, 264766, 56181562, 18108359, 18108365, 18108370, 18108381
660	81897822 (1319, 1320)	Novel Protein sim. GBank gi 134180 sp P15401 SACY_BACSU - LEVANSUCRASE AND SUCRASE SYNTHESIS OPERON ANTITERMINATOR	Contains protein domain (PF00874) - Transcriptional antiterminaltor bglG family	UNCLASSIFIED	264757
662	20463731 (1323, 1324)	Novel Protein sim. GBank gi 4545229 gb AAD22450.1 AF11618 - (AF116183) SecA homolog [Actinobacillus actinomycetemcomitans]		UNCLASSIFIED	264605
663	20628080 (1325, 1326)	Novel Protein sim. GBank gi 5689250 dbj BAA02881.1  - (AB024335) similar to orf5 [Comamonas testosteroni]		dehydrogenase	264605
664	80508512 (1327, 1328)	Novel Protein sim. GBank gi 1652848 dbj BAA17766 - (D90909) DNA photolyase [Synechocystis sp.]		UNCLASSIFIED	264769
665	80079053 (1329, 1330)	Novel Protein sim. GBank gi 116841 sp P21640 COBJ_PSEDE - PRECORRIN-38 C17 METHYLTRANSFERASE (PRECORRIN-3 METHYLTRANSFERASE)		isomerase	264600
668	79603142 (1331, 1332)	Novel Protein sim. GBank gi 3261829 emb CAB109271 - (Z98260) hypothetical protein Rv1230c [Mycobacterium tuberculosis]		glycoprotein	264907, 265007
667	94631802 (1333, 1334)	Novel Protein sim. GBank gi 5688851 dbj BAA82702.1  - (AB017438) Orf5 [Streptomyces coelicolor]		UNCLASSIFIED	264689, 264602, 264593
668	82051891 (1335, 1336)	Novel Protein sim. GBank gi 3581853 emb CAA20809 - (AL031541) 50S ribosomal protein L20 [Streptomyces coelicolor]	Contains protein domain (PF00493) - Ribosomal protein L20	ribosomalprot	264905, 264908, 264908, 264600, 264601, 264603, 264805, 264760, 264689, 264636, 264638, 264639
669	12867154 (1337, 1338)	Novel Protein sim. GBank gi 2482531 (AF026444) - 2- isopropylmalate synthase [Streptomyces coelicolor]		UNCLASSIFIED	264637
670	80238548 (1339, 1340)	Novel Protein sim. GBank gi 2482531 (AF026444) - 2- isopropylmalate synthase [Streptomyces coelicolor]		synthase	264905, 264906, 264908, 264601, 264762, 264766, 264689, 264638, 18108385, 264488
671	79601358 (1341, 1342)	Novel Protein sim. GBank gi 2114430 (U92703) - Olf-1/EBF-like-3 transcription factor [Mus musculus]	Contains protein domain (PF00023) - Ank repeat	UNCLASSIFIED	264690, 264692, 264693, 264636, 18108387
672	79834371 (1343, 1344)	Novel Protein sim. GBank gi 2114430 (U92703) - Olf-1/EBF-like-3 transcription factor [Mus musculus]		transcriptfactor	264810, 265017
673	82283788 (1345, 1346)	Novel Protein sim. GBank gi 4589285 gb AAD26430.1 AF13515 - (AF135154) ferric alicillin siderophore receptor [Bordetella pertussis]			264759
674	78189259 (1347, 1348)			UNCLASSIFIED	264629

675	87895870 (1349, 1350)	Novel Protein sim. GBank gi14980755 p p A D 35347.1 AE00170 - (AE001708) D- alanine--D-alanine ligase [Thermotoga maritima]	Contains protein domain (PF01820) - D-alta D-ala ligase	UNCLASSIFIED	264488, 22278989, 86714117, 264508, 204511, 285008, 60433438, 264600, 264601, 264602, 264603, 264604, 264605, 264762, 264687, 264769, 60431602, 18108374, 264636, 264638 285010
676	78895607 (1351, 1352)	Novel Protein sim. GBank gi1723566 p Q10479 VDF7_SCHPO - PUTATIVE GLUCOSYLTRANSFERASE C17C9.07			
677	21644312 (1353, 1354)	Novel Protein sim. GBank gi587208 (U03976) - dynein heavy chain IsoType 5C [Tritoneustes gratilla]		ATPase-associated	264591, 264632
678	84225200 (1355, 1356)	Novel Protein sim. GBank gi1588274 p rfj 2203365A - laminin alpha5 [Mus musculus]	Contains protein domain (PF00053) - Laminin EGF-like (Domains III and V)	laminin	264758, 264682, 264657
679	78868855 (1357, 1358)	Novel Protein sim. GBank gi13928723 emb CAA22219  - (AL034355) putative ABC transporter [Streptomyces coelicolor]		UNCLASSIFIED	22278986, 264693
680	20726424 (1359, 1360)				
681	84322017 (1361, 1362)	Novel Protein sim. GBank gi15174483 ref NP_006050.1 p LAMC - laminin, gamma 3	Contains protein domain (PF00053) - Laminin EGF-like (Domains III and V)	laminin	264600, 264602 264102, 264907, 264808, 265006, 264693, 263972, 83373044, 264668
682	11392476 (1363, 1364)				
683	80083680 (1365, 1366)	Novel Protein sim. GBank gi14758208 ref NP_004081.1 p DUSP - dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	Contains protein domain (PF00782) - Dual specificity phosphatase, catalytic domain	UNCLASSIFIED phosphatase	284595 264634
684	20465387 (1367, 1368)	Novel Protein sim. GBank gi15420387 emb CAB46679.1  - (AJ243459) proteophosphoglycan [Leishmania major]			284605
685	80248735 (1369, 1370)				
686	78208608 (1371, 1372)				
687	80085629 (1373, 1374)				
688	78853412 (1375, 1376)	Novel Protein sim. GBank gi12688962 (AF027768) - LspA [Serratia marcescens]		peptidase	264809, 263987, 263981 264631 264693, 264635 264907, 264638
689	88064256 (1377, 1378)	Novel Protein sim. GBank gi13046931 (AF049330) - PPAR gamma coactivator [Mus musculus]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264906, 264907, 265007, 265009, 60433438, 21908754, 264760, 18108358, 21908768, 21908769, 265021, 18108381, 263974, 18108379, 264557, 18108385, 22279002 264510, 264511, 264764, 264769
690	80389750 (1379, 1380)	Novel Protein sim. GBank gi1248894 p p Q15428 SP62_HUMAN - SPICEOSOME ASSOCIATED PROTEIN 92 (SAP 92) (SF3A66)			
691	81654392 (1381, 1382)			UNCLASSIFIED	264757
692	83608936 (1383, 1384)	Novel Protein sim. GBank gi15420387 emb CAB46678.1  - (AJ243459) proteophosphoglycan [Leishmania major]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	55812038, 55811937, 265018, 55811150, 18108351, 264908, 60431528, 264594
693	79586116 (1385, 1386)	Novel Protein sim. GBank gi1854065 emb CAA583371 - (X83413) U88 [Human herpesvirus 6]			264635
694	82455983 (1387, 1388)	Novel Protein sim. GBank gi1267327 p Q101033 VG48_HSVSA - HYPOTHETICAL GENE 48 PROTEIN			22278996, 264510, 264602, 264603, 264762, 264687, 264769, 264693

695	84147849 (1389, 1390)	Novel Protein sim. GBank gij4468339[emb]CAB38059.1[- (AJ010901) MUC4 (Homo sapiens)]	Contains protein domain (PF00094) - von Willebrand factor type D domain	UNCLASSIFIED	56182375, 264509, 264805, 264907, 29331630, 264908, 264909, 264511, 265007, 264910, 264758, 264764, 264288, 65274791, 264905, 264595
696	78830982 (1391, 1392)	Novel Protein sim. GBank gij2649950 (AE001058) - glutamine ABC transporter, ATP-binding protein (ginQ) [Archaeoglobus fulgidus]	Contains protein domain (PF00005) - transport ABC transporter		
697	11767869 (1393, 1394)	Novel Protein sim. GBank gij1731343[sp]Q10694[YY25_MYCTU - HYPOTHETICAL 24.4 KD PROTEIN CY49.25]	Contains protein domain (PF01836) - Transposase	UNCLASSIFIED	264882
698	86695862 (1395, 1396)			UNCLASSIFIED	264688, 35695917
699	79582558 (1397, 1398)			UNCLASSIFIED	264682
700	79639098 (1399, 1400)				264693
701	80230242 (1401, 1402)	Novel Protein sim. GBank gij1001236[db]BAA10471[- (DB4003) Hypothetical protein [Synchocystis sp.]		UNCLASSIFIED	264488, 264510, 264511, 264602, 264605, 264689
702	79814788 (1403, 1404)			UNCLASSIFIED	264809
703	20446820 (1405, 1406)	Novel Protein sim. GBank gij2498935[sp]Q46338[SOXG_CORSP - SARCOSINE OXIDASE GAMMA SUBUNIT]	oxidase		264804
704	94312224 (1407, 1408)	Novel Protein sim. GBank gij3150513 (AF067219) - contains similarity to the kelch/MIPP family [Caenorhabditis elegans]	Contains protein domain (PF01344) - Kelch motif	UNCLASSIFIED	264288, 56181562, 33657109, 264629, 55811576
705	17632141 (1409, 1410)	Novel Protein sim. GBank gij421091[pir]S30730 - hypothetical protein o206 - Escherichia coli		UNCLASSIFIED	265006
706	20286082 (1411, 1412)	Novel Protein sim. GBank gij3024872[sp]Q55790[Y074_SNNY3 - HYPOTHETICAL 52.8 KD PROTEIN SLR0074]			264600
707	20638065 (1413, 1414)	Novel Protein sim. GBank gij3420608[gb]AAC31907.1[- (AF075709) ABC transporter ATP-binding subunit [Pseudomonas putida]	transport		264603
708	20702892 (1415, 1416)	Novel Protein sim. GBank gij3649741[emb]CAA03985[- (AJ000281) mucin [Homo sapiens]			264801, 264692
709	88001438 (1417, 1418)	Novel Protein sim. GBank gij3080425[emb]CAA18744.1[- (AL022804) putative protein [Arabidopsis thaliana]	slrud		18108398, 264637, 264908, 264909
710	11356683 (1419, 1420)				264369
711	17931418 (1421, 1422)	Novel Protein sim. GBank gij1703266[sp]Q11056[AM12_MYCTU - PUTATIVE AMIDASE CY50.19C]			265019
712	80258164 (1423, 1424)	Novel Protein sim. GBank gij4758686[ref]NP_002323.1[plLRP1 - low density lipoprotein related protein 1 (alpha-2-macroglobulin receptor)]	Contains protein domain (PF00058) - Low-density lipoprotein receptor repeat class B	apolipoprotein	264591
713	79263126 (1425, 1426)	Novel Protein sim. GBank gij1703266[sp]Q11056[AM12_MYCTU - PUTATIVE AMIDASE CY50.19C]		hydrolase	264906, 264907
714	27847651 (1427, 1428)	Novel Protein sim. GBank gij4502351[ref]NP_001692.1[pBAAT - bile acid Coenzyme A: amino acid N-acetyltransferase, glycine N-choleyltransferase]			264508, 264555

715	78639423 (1428, 1430)	Novel Protein sim. GBank gii1789035 (AE000352) - ori, hypothetical protein [Escherichia coli]		UNCLASSIFIED	264907
716	79555072 (1431, 1432)				264992
717	79491842 (1433, 1434)	Novel Protein sim. GBank gii2494074 (spIP55653) (GABD_RHISN - PROBABLE SUCCINATE-SEMIALDEHYDE DEHYDROGENASE [NADP+]) (SSDH)	dehydrogenase		264538
718	94315658 (1435, 1436)	Novel Protein sim. GBank gii3673579 (emb) (CAA94886) - (Z71178) similar to pre-collagen domains; cDNA EST EMBL:D27978 comes from this gene; cDNA EST EMBL:D27977 comes from this gene; cDNA EST EMBL:D34199 comes from this gene; cDNA EST EMBL:D84392 comes from this gene; cDNA EST EMBL... (Z95387) hypothetical protein Rv2811c [Mycobacterium tuberculosis]	kinase		18108392, 22278994, 22278998, 265008, 265018, 264681, 18108354, 264684, 264685, 264686, 264687, 264689, 21906769, 18108361, 264691, 264692, 55810764, 264635, 18108381, 18108382, 83373044, 18108388
718	17679584 (1437, 1438)	Novel Protein sim. GBank gii2104302 (emb) (CAB08631) - (Z95387) hypothetical protein Rv2811c [Mycobacterium tuberculosis]	UNCLASSIFIED		265011
720	78841684 (1438, 1440)	Novel Protein sim. GBank gii123530 (spIP04929) (HRPX_PLALO - HISTIDINE-RICH GLYCOPROTEIN PRECURSOR	UNCLASSIFIED		264908
721	15020180 (1441, 1442)	Novel Protein sim. GBank gii498253 (U02372) - integrase [Vibrio cholerae]			264629
722	8862603 (1443, 1444)	Novel Protein sim. GBank gii2253054 (emb) (CAB10705) - (Z97559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]	UNCLASSIFIED		264910
723	19755599 (1445, 1446)	Novel Protein sim. GBank gii4063015 (AF083061) - protease PrtA [Pseudomonas fluorescens]	protease		264681
724	10126494 (1447, 1448)		Contains protein domain (PF00353) - Hemolysin-type calcium-binding proteins		264909
725	78878579 (1449, 1450)			UNCLASSIFIED	264905, 264907
726	13086282 (1451, 1452)			UNCLASSIFIED	264636
727	13522872 (1453, 1454)				264634
728	20268471 (1455, 1456)	Novel Protein sim. GBank gii2833910 (emb) (CAB13411) - (Z99112) similar to hypothetical proteins [Bacillus subtilis]			264567
729	11293753 (1457, 1458)	Novel Protein sim. GBank gii2494660 (spIQ45291) (GALE_BRELA - UDP-GLUCOSE 4-EPIMERASE (GALACTOWALDENASE) (UDP-GALACTOSE 4-EPIMERASE)	isomerase		264490
730	19500373 (1459, 1460)	Novel Protein sim. GBank gii1146192 (L47838) - putative [Bacillus subtilis]	UNCLASSIFIED		264564
731	80058750 (1461, 1462)	Novel Protein sim. GBank gii1168396 (spIP46881) (AIP2_YEAST - ACTIN INTERACTING PROTEIN 2	UNCLASSIFIED		264605
732	80258175 (1463, 1464)		struct		264591, 264594, 264595
733	20446839 (1465, 1466)	Novel Protein sim. GBank gii3184080 (emb) (CAA19336) - (AL023781) hypothetical protein [Schizosaccharomyces pombe]	UNCLASSIFIED		264604
734	20435987 (1467, 1468)		ubiquitin		264604

735	11607959 (1469, 1470)	Novel Protein sim. GBank gi 401582 sp P27432 YICE_ECOLI - HYPOTHETICAL 48.9 KD PROTEIN IN GLTS-SELC INTERGENIC REGION				264594
736	10878734 (1471, 1472)	Novel Protein sim. GBank gi 400831 sp P31135 POTH_ECOLI - PUTRESCINE TRANSPORT SYSTEM PERMEASE PROTEIN POTH	Contains protein domain (PF00528) - Binding-protein-dependent transport systems inner membrane component	transport		264636
737	78945340 (1473, 1474)		Contains protein domain (PF00815) - Regulator of G protein signaling domain	UNCLASSIFIED		265020
738	17895353 (1475, 1476)	Novel Protein sim. GBank		oxidase		265008
739	79833870 (1477, 1478)	gi 2506867 sp P33225 TORA_ECOLI - TRIMETHYLAMINE- N-OXIDE REDUCTASE PRECURSOR (TMAO REDUCTASE) (TRIMETHYLAMINE OXIDASE)				264910
740	19881557 (1479, 1480)	Novel Protein sim. GBank gi 3261828 emb CA810925  - (Z88260) mmp [Mycobacterium tuberculosis]	Contains protein domain (PF01883) - Domain of unknown function	UNCLASSIFIED		264907, 264764, 264634, 264637
741	79827273 (1481, 1482)	Novel Protein sim. GBank gi 3877494 emb CAA88472.1  - (Z48583) ATP binding protein with similarity to the CDC48/PAS1/SEC18 family; cDNA EST EMBL:D65037 comes from this gene; cDNA EST EMBL:D68340 comes from this gene; cDNA EST EMBL:D65048 comes from this gene; cDNA EST EMBL:D6845...	Domain of unknown function	UNCLASSIFIED		264688, 35696286, 264510, 264908, 18108362
742	82393795 (1483, 1484)	Novel Protein sim. GBank gi 127420 sp P19888 MTBA_BACAR - MODIFICATION METHYLASE BANI (CYTOSINE-SPECIFIC METHYLTRANSFERASE BANI) (M.BANI)	Contains protein domain (PF00145) - C-5 cytosine-specific DNA methylase	UNCLASSIFIED		264488, 264259, 264508, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 265008, 265009, 264910, 264591, 264596, 264759, 265010, 265011, 18108351, 264763, 264288, 264766, 264768, 264693, 18108370, 264629, 18108372, 264630, 264631, 264634, 264556, 18108385, 264482, 264564, 264567
743	82300051 (1485, 1486)	Novel Protein sim. GBank gi 127420 sp P19888 MTBA_BACAR - MODIFICATION METHYLASE BANI (CYTOSINE-SPECIFIC METHYLTRANSFERASE BANI) (M.BANI)				18108397, 264511, 264690, 264628, 264638, 264692, 264639, 264768
744	80230421 (1487, 1488)			glycoprotein		264906
745	9841963 (1489, 1490)	Novel Protein sim. GBank gi 78921 pr I SO4846 - UDP-N- acetyluracilamoyl-D-glutaryl-2, 6-diaminopimelate-D- alanine-D-alanine ligase (EC 6.3.2.15) precursor - Escherichia coli				
746	11073229 (1491, 1492)	Novel Protein sim. GBank gi 3386354 (AF074705) - pyochelin synthetase [Pseudomonas aeruginosa]		synthase		264600
747	94322044 (1493, 1494)	Novel Protein sim. GBank gi 2887411 dbj BA24848  - (AB007878) KIAA0418 [Homo sapiens]	Contains protein domain (PF00018) - SH3 domain	oxidase		66714117, 264905, 264509, 264906, 264907, 264908, 264909, 264511, 264910, 265011, 264681, 264288, 264766, 264687, 264768, 264769, 21906768, 35695917, 264691, 264693, 264628, 264634, 264635, 264639, 58182323, 83373044
748	11617923 (1495, 1496)					264690

749	20499118 (1497, 1498)	Novel Protein sim. GBank gij1169727spj44948jFPG_HAEIN - FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE (FAPY- DNA GLYCOSYLASE)		UNCLASSIFIED	284604 284600
750	20296427 (1489, 1500)				
751	21636169 (1501, 1502)	Novel Protein sim. GBank gij5360088jgbjAA0x2851.1jAF15968 - (AF15968) serine/threonine kinase PKM3 [Myxococcus xanthus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	284605, 284559
752	82450366 (1503, 1504)	Novel Protein sim. GBank gij1168682spj44426jBIOA_HAEIN - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7.8-DIAMINO-PELARGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)	Contains protein domain (PF00202) - Aminotransferases class-III pyridoxal phosphate	UNCLASSIFIED	284508, 284907, 284510, 265011, 284762, 284698, 35899855, 284638, 18108387
753	80508718 (1505, 1506)	Novel Protein sim. GBank gij2851530spj232399jYHGE_BACSU - HYPOTHETICAL 84.1 KD PROTEIN IN HEMY-GLTT INTERGENIC REGION (ORFB)		UNCLASSIFIED	284909, 284600, 284602, 284604, 284760, 284789, 284634
754	95083741 (1507, 1508)				
755	80185449 (1509, 1510)				
756	94631686 (1511, 1512)	Novel Protein sim. GBank gij3449276jembjCAA20420j - (AL031317) putative dehydrogenase [Streptomyces coelicolor]		UNCLASSIFIED	284508, 284905, 284907, 284908, 284909, 284759, 284602, 284764, 284789, 284828, 284629, 284630, 284632, 284634, 284635, 284637, 284638, 83373044, 18108385
757	79468533 (1513, 1514)	Novel Protein sim. GBank gij4580331jembjCAB40107.1j - (AJ001206) putative glycogen debranching enzyme [Streptomyces coelicolor]		UNCLASSIFIED	284448, 284690
758	78863176 (1515, 1516)			UNCLASSIFIED	284768, 284689, 284638, 284639
759	79475667 (1517, 1518)	Novel Protein sim. GBank gij2811858 (AF047659) - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	284682, 284685
760	87628888 (1519, 1520)	Novel Protein sim. GBank gij3451312jembjCAA20449j - (AL031324) membrane alase [Schizosaccharomyces pombe]	Contains protein domain (PF00122) - E1-E2 ATPase	amylase	265007, 18108387, 265007, 18108387
761	79877966 (1521, 1522)			UNCLASSIFIED	284684, 284688
762	80023563 (1523, 1524)	Novel Protein sim. GBank gij3327158jbjBAA31547j - (AB014572) KIAA0572 protein [Homo sapiens]	Contains protein domain (PF00122) - E1-E2 ATPase	transport	29331822, 284908, 52844045, 56182435, 60170831, 21906754, 285017, 285019, 284681, 284687, 284688, 21806786, 21906788, 265020, 265021, 265022, 284835, 22275000
783	20284813 (1525, 1526)	Novel Protein sim. GBank gij4981266jbjAAD35822.1jAE00174 - (AE001744) lipopolysaccharide core biosynthesis protein KdIB [Thermotoga maritima]		UNCLASSIFIED	264768
764	39515024 (1527, 1528)			UNCLASSIFIED	284907, 284593, 265020
					284600
					284603

765	80025347 (1528, 1530)	Novel Protein sim. GBank gij3845093 (AE001371) - erythrocyte membrane protein P(EMP3) [Plasmodium falciparum]			264905, 264805, 264594, 264686, 33657023
766	82417404 (1531, 1532)				
767	10296742 (1533, 1534)	Novel Protein sim. GBank gij541121 [pril] [S40827] - hypothetical protein o300 - Escherichia coli		UNCLASSIFIED	264805, 284762, 18108374
768	79416080 (1535, 1536)				264907
769	80086554 (1537, 1538)	Novel Protein sim. GBank gij2982501 [emb] [CAA06164] - (AJ004832) neuropathy target esterase [Homo sapiens]		UNCLASSIFIED	264592, 264595
770	80417647 (1539, 1540)	Novel Protein sim. GBank gij283437 [pril] [S27850] - hypothetical protein - Trypanosoma cruzi (fragment)		esterase	55810764, 264559
771	95329509 (1541, 1542)	Novel Protein sim. GBank gij4768004 [gbl] [AD29715.1] [AF14059] - ring-box protein 1 [Homo sapiens]	Contains protein domain (PF000097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	264905, 264907, 264828, 264909, 265010, 264766, 264628, 264629, 264634, 264636, 264555
772	78971362 (1543, 1544)				
773	78945363 (1545, 1546)				
774	79856128 (1547, 1548)	Novel Protein sim. GBank gij5531324 [emb] [CAB51045.1] - (AJ009579) putative alkane 1-monooxygenase [Pseudomonas fluorescens]		UNCLASSIFIED	265020
775	20620141 (1549, 1550)				264909
776	78942693 (1551, 1552)				
777	79960378 (1553, 1554)	Novel Protein sim. GBank gij4505461 [re] [NP_003624.1] [pNRPB] - nuclear restricted protein, BTB domain-like (brain)	Contains protein domain (PF01006) - Hepatitis C virus non-structural protein NS4a	UNCLASSIFIED	264555
778	20691310 (1555, 1556)				285019
779	80054024 (1557, 1558)				
780	95288987 (1559, 1560)	Novel Protein sim. GBank gij1144320 (U34856) - phosphoribosylformylglycinamide synthase [Mycobacterium tuberculosis]	Contains protein domain (PF01344) - Ketch motif	protease	21908754, 265020, 60170615, 264691
781	80250049 (1561, 1562)				
782	8766528 (1563, 1564)	Novel Protein sim. GBank gij415447 (AE001517) - proline/betaine transporter [Helicobacter pylori J99]		UNCLASSIFIED	264511
783	16410791 (1565, 1566)				264603
784	80051197 (1567, 1568)				
785	56073541 (1569, 1570)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]		synthase	264907, 264600, 264601, 264602, 264603, 264604, 264605, 264486
786	20438842 (1571, 1572)	Novel Protein sim. GBank gij138748 [sp] [P10905] [UGPA_ECOLI] - SN-GLYCEROL-3-PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN UGPA	Contains protein domain (PF00047) - Immunoglobulin domain	transport	264905, 264907, 265010, 264600, 264601, 18108382, 18108374, 264556
787	80258364 (1573, 1574)				264605
					265020
					264635, 33657023, 29331828, 265017, 264585, 264586
					35989052, 264604
					264603
					264593



788	80507844 (1575, 1576)	Novel Protein sim. GBank gij2748079 (AF015310) - BTH1 [Brassica napus]			synthase	264809, 264802, 264603, 264769, 264638
789	17294715 (1577, 1578)	Novel Protein sim. GBank gij2351849 (U93357) - 40 kDa heat shock chaperone protein [Halobacterium salinarum]			UNCLASSIFIED	265007
790	86284406 (1579, 1580)	Novel Protein sim. GBank gij5708378 [dij]BAA63099.1] - (AB026118) MALT1 [Homo sapiens]		Contains protein domain (PF00047) - Immunoglobulin domain	glycoprotein	22278999, 29331824, 264828, 87168559, 265018, 21906765, 21906767, 21908768, 21908769, 265020, 264892, 22278000, 264563
791	94851627 (1581, 1582)	Novel Protein sim. GBank gij5689948 [emb]CAB51985.1] - (AL109663) putative Isoleucyl-tRNA synthetase [Streptomyces coelicolor A3(2)]				264601, 264605, 264636
792	80055786 (1583, 1584)	Novel Protein sim. GBank gij393194 (L02375) - S-anilgen [Plasmodium falciparum]			struct	265021, 264831, 264635, 264656
793	79838730 (1585, 1586)	Novel Protein sim. GBank gij1345408 [db]BAA05046] - (D28046) AT motif-binding factor [Mus musculus]		Contains protein domain (PF00046) - Homeobox domain	homeobox	264693
794	81839284 (1587, 1588)	Novel Protein sim. GBank gij105884 [prij]S24023 - dopamine receptor D4 - human (fragment)			UNCLASSIFIED	264603, 264604, 264910, 264762, 264908, 264639, 264909, 264757
795	80074988 (1589, 1590)	Novel Protein sim. GBank gij1877334 [emb]CAB07082] - (Z92771) bira [Mycobacterium tuberculosis]			carboxylase	264488, 35696052, 264905, 264907, 265010, 35696423, 264636
796	86669451 (1591, 1592)	Novel Protein sim. GBank gij2995447 [emb]CAA71519] - (Y10485) CDV-1R protein [Mus musculus]				60432229, 55811150, 264630, 264637, 264565
797	87771781 (1593, 1594)	Novel Protein sim. GBank gij4467250 [emb]CAB37575] - (AL035569) probable Glu-tRNA Gln amidotransferase subunit [Streptomyces coelicolor]			struct	22278998, 264093, 264094, 66714117, 21908767, 21908769, 265020, 265022
798	79865209 (1595, 1596)	Novel Protein sim. GBank gij1467250 [emb]CAB37575] - (AL035569) probable Glu-tRNA Gln amidotransferase subunit [Streptomyces coelicolor]			transcriptfactor	264687, 264768, 264693
799	79557816 (1597, 1598)	Novel Protein sim. GBank gij1467250 [emb]CAB37575] - (AL035569) probable Glu-tRNA Gln amidotransferase subunit [Streptomyces coelicolor]			hydrolase	264909, 264910, 264638, 264638
800	79870189 (1599, 1600)	Novel Protein sim. GBank gij2791517 [emb]CAA16054] - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]			UNCLASSIFIED	264488
801	80499389 (1601, 1602)	Novel Protein sim. GBank gij2791517 [emb]CAA16054] - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]			transport	264508, 264511, 265008, 265009, 264789, 264567, 264486
802	79834598 (1603, 1604)	Novel Protein sim. GBank gij4487211 [gb]AAD32237.1 [AF14744 - (AF147449) penicillin binding protein 1B [Pseudomonas aeruginosa]				264905, 264693
803	20467520 (1605, 1606)	Novel Protein sim. GBank gij1176152 [sp]P44507 [YHAD_HAEIN - HYPOTHETICAL PROTEIN H0091]			struct	264605
804	10174239 (1607, 1608)	Novel Protein sim. GBank gij1176152 [sp]P44507 [YHAD_HAEIN - HYPOTHETICAL PROTEIN H0091]			kinase	264510
805	79599993 (1609, 1610)	Novel Protein sim. GBank gij2764612 [emb]CAA04683] - (AJ001330) ornithine transcarbamoylase [Lactobacillus sakei]		Contains protein domain (PF00185) - Aspartate/ornithine carbamoyltransferase		264508
806	80484113 (1611, 1612)	Novel Protein sim. GBank gij2764612 [emb]CAA04683] - (AJ001330) ornithine transcarbamoylase [Lactobacillus sakei]			transferase	264769
807	80381812 (1613, 1614)	Novel Protein sim. GBank gij283331 [sp]Q21828 [YNF0_CAEEL - HYPOTHETICAL 18.9 KD PROTEIN R07ES.13 IN CHROMOSOME III]				264764

808	35108817 (1615, 1616)	Novel Protein sim. GBank gi3913092[sp]Q46170[ARCD_CLOPE - ARGININE/ORNITHINE ANTIPORTER		transport	264909, 264602, 21906764, 16108374
809	81454254 (1617, 1618)	Novel Protein sim. GBank gi3913016[sp]P74309[ALF1_SYN3 - FRUCTOSE- BISPHOSPHATE ALDOLASE CLASS (FBP ALDOLASE)	Contains protein domain (PF00274) - Fructose-bisphosphate aldolase class-I	UNCLASSIFIED	264508, 264906, 264909, 265007, 264910, 264758, 264600, 264602, 264603, 264605, 264687, 264769, 264689, 264636, 264488
810	80192781 (1618, 1620)	Novel Protein sim. GBank gi401472[sp]P30863[YAFB_ECOLI - HYPOTHETICAL OXIDOREDUCTASE IN ASPU-MLTD INTERGENIC REGION	Contains protein domain (PF00248) - Aldoketo reductase family	reductase	264369
811	80079280 (1621, 1622)			UNCLASSIFIED	264558
812	10297654 (1623, 1624)			UNCLASSIFIED	264692
813	79612280 (1625, 1626)			UNCLASSIFIED	264808
814	80473427 (1627, 1628)	Novel Protein sim. GBank gi146168 (J01617) - glutaminyl- [RNA synthetase [Escherichia coli]		synthase	264805, 264602, 264605, 264882, 264687, 264769, 264636
815	95419513 (1628, 1630)	Novel Protein sim. GBank gi4589652[dbj]BAA76848.1 - (A8023221) KIAA1004 protein [Homo sapiens]		UNCLASSIFIED	264488, 22278998, 22278999, 29331822, 29331824, 29331825, 29331827, 29331828, 29146499, 264905, 264908, 265007, 33657402, 60433356, 60433438, 264758, 265011, 265017, 265018, 265019, 264388, 264288, 264685, 21906765, 21906767, 265020, 265021, 264692, 65274620, 33657109, 264629, 18108376, 264635, 264638, 60170394, 56182323, 264564, 264600
816	19881910 (1631, 1632)	Novel Protein sim. GBank gi1781144[emb]CAB06254 - (Z83866) hypothetical protein Rv3068 [Mycobacterium tuberculosis]		UNCLASSIFIED	264595
817	95293316 (1633, 1634)	Novel Protein sim. GBank gi1477468 (J35244) - vacuolar protein sorting homolog v-ps33a [Rattus norvegicus]			65274572, 22278998, 60424269, 35696032, 55812038, 21906768, 55811957, 35695917, 33657023, 18108370, 18108374, 55810764, 35696423, 55811576, 264636
818	80938180 (1635, 1636)			transport	264565
819	80254977 (1637, 1638)	Novel Protein sim. GBank gi1001352[dbj]BAA10839 - (D64008) ABC transporter [Synecchocystis sp.]		UNCLASSIFIED	264600, 264602, 264804
820	80059688 (1639, 1640)	Novel Protein sim. GBank gi586814[sp]P37484[YVT_BACSU - HYPOTHETICAL 74.3 KD PROTEIN IN RPL1-COTF INTERGENIC REGION			
821	79782590 (1641, 1642)			UNCLASSIFIED	264910
822	80215310 (1643, 1644)	Novel Protein sim. GBank gi3878400[emb]CAA95828 - (Z71264) predicted using GeneFinder. Weak similarity to Mouse T-complex-associated-lactes-expressed-1 protein (PIR Acc. No. A45841); cDNA EST EMBL:D32742 comes from this gene; cDNA EST EMBL:D33617 comes from this gene; cDNA EST...		struct	264510, 264594, 264637, 264509, 264687, 264691
823	94992299 (1645, 1646)				
824	80411171 (1647, 1648)	Novel Protein sim. GBank gi1370075[emb]CAA66887 - (X98235) type I [Drosophila melanogaster]	Contains protein domain (PF01429) - Methyl-CpG binding domain		264910, 264763, 264769, 264593

825	20638600 (1649, 1650)	Novel Protein sim. GBank gi 3025132 sp P77391 YEAG_ECOLI - HYPOTHETICAL 74.5 KD PROTEIN IN GAPA-RND INTERGENIC REGION		UNCLASSIFIED	264592
826	11075047 (1651, 1652)	Novel Protein sim. GBank gi 3242281 emb CAA16659  - (AL021646) hypothetical protein Rv3202c [Mycobacterium tuberculosis]			264605
827	80054207 (1653, 1654)	Novel Protein sim. GBank gi 3417424 emb CAA20312  - (AL031261) putative transport protein (Schizosaccharomyces pombe)			264603
828	95106322 (1655, 1656)	Novel Protein sim. GBank gi 4336602 gb AA017897  - (AF101381) Abnormal X segregation [Drosophila melanogaster]		UNCLASSIFIED	52845080, 264508, 264805, 264805, 264509, 264906, 264907, 264908, 264909, 264512, 264910, 264591, 264758, 264800, 264766, 264768, 21906768, 35695917, 264691, 264828, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264658, 264639, 83373044, 18108385, 264563, 264568, 264486
829	81742215 (1657, 1658)	Novel Protein sim. GBank gi 3820539 (AF080002) - UDP-N- acetyluracetyl tripeptide synthetase MurC [Helicobacter modulus]		UNCLASSIFIED	264758, 264634
830	20396091 (1659, 1660)			UNCLASSIFIED	264603
831	87112435 (1661, 1662)	Novel Protein sim. GBank gi 1870004 emb CA068655  - (Z92559) hypothetical protein Rv1024 [Mycobacterium tuberculosis]		UNCLASSIFIED	66714117, 264910, 264639
832	18536322 (1663, 1664)			UNCLASSIFIED	264906
833	20726654 (1665, 1666)	Novel Protein sim. GBank gi 2500055 sp Q46267 PFLA_CLOPA - PYRUVATE FORMATE-LYASE ACTIVATING ENZYME		UNCLASSIFIED	264802
834	21428762 (1667, 1668)	Novel Protein sim. GBank gi 2497531 sp Q46078 PKYK_CORGL - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate kinase	UNCLASSIFIED	264800, 264802, 264769, 264889, 264836
835	94140482 (1669, 1670)				264768, 263994, 21906767, 264910, 264632, 264635, 264258, 264639, 264693, 83373044, 264758, 35696052, 22279002, 264508, 264905, 264906, 264448, 263972, 264908, 264909
836	68126552 (1671, 1672)	Novel Protein sim. GBank gi 699315 (U15184) - phosphate transport protein PSTC [Mycobacterium leprae]		transport	35695917, 264557
837	79450450 (1673, 1674)	Novel Protein sim. GBank		UNCLASSIFIED	264595
838	78184203 (1675, 1676)	gi 728867 sp P40602 APG_ARATH - ANTER-SPECIFIC PROLINE-RICH PROTEIN APG PRECURSOR		UNCLASSIFIED	264687
839	79641125 (1677, 1678)	Novel Protein sim. GBank gi 2495533 sp Q50598 Y0D8_MYCTU - HYPOTHETICAL 69.9 KD PROTEIN CV1A11.08		UNCLASSIFIED	264906
840	80056851 (1679, 1680)	Novel Protein sim. GBank gi 4557753 sp NP_000372.1 pMD1 - midline 1 protein (finer)	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	intra leukinrecept	264762, 264558

841	80376318 (1681, 1682)	Novel Protein sim. GBank gll139805 sp P08045 XFN_XENLA - XFN PROTEIN	Contains protein domain (PF00006) - Zinc finger, C2H2 type	transcript factor	264764
842	80078724 (1683, 1684)	Novel Protein sim. GBank gll2114321 dbj BAA200371 - (D88733) membrane glycoprotein [Equine herpesvirus 1]	Contains protein domain (PF00569) - Zinc finger present in dystrophin, CBP/p300	UNCLASSIFIED	264805, 264908, 265008, 265009, 18108374, 58182323, 264558
843	87002847 (1685, 1686)	Novel Protein sim. GBank gll3882325 dbj BAA34522.1 - (AB018345) KIAA0802 protein [Homo sapiens]	Contains protein domain (PF00170) - bZIP transcription factor	struct	264091, 29331825, 264906, 264768, 264563
844	17941439 (1687, 1688)	Novel Protein sim. GBank gll2224721 dbj BAA20844 - (AB002388) KIAA0390 [Homo sapiens]	Contains protein domain (PF00066) - Zinc finger, C2H2 type	transcript factor	265011
845	18346844 (1689, 1690)				264829
846	79863441 (1691, 1692)	Novel Protein sim. GBank gll625679 pir A36929 - virulence regulatory protein VsrB - Pseudomonas solanacearum		kinase	264907
847	78693348 (1693, 1694)				264809
848	78489385 (1695, 1696)			UNCLASSIFIED	265020
849	79756387 (1697, 1698)			UNCLASSIFIED	264566
850	78817849 (1699, 1700)	Novel Protein sim. GBank gll3183245 sp P78061 YCKJ_ECOLI - PUTATIVE GLUTAMINE SYNTHETASE (GLUTAMATE-AMMONIA LIGASE)	Contains protein domain (PF00120) - Glutamine synthetase	UNCLASSIFIED	264809
851	95320333 (1701, 1702)	Novel Protein sim. GBank gll5454130 ref NP_006280.1 pt TLN - talin	Contains protein domain (PF01608) - ILWEEQ domain		264488, 52644507, 264489, 18108388, 65274572, 56182575, 22278994, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 20281171, 264490, 264259, 52645080, 29331822, 29331824, 66714117, 29331825, 60432289, 29331826, 29331827, 35696052, 29331828, 29146498, 29146499, 264107, 264905, 264906, 264907, 264908, 52644045, 56182435, 265006, 265007, 265008, 265009, 284910, 80432229, 60431735, 60433356, 33657402, 60433438, 264595, 264758, 264759, 21906754, 33108954, 52644296, 265010, 265011, 87168559, 285017, 265018, 265019, 264760, 264761, 264762, 264681, 18108351, 264763, 264448, 264882, 264764, 264883, 18108354, 284288, 284389, 264885, 264768, 264887, 264768, 264769, 21906765, 21906766, 21906767, 21906768, 29148627, 21906769, 29148629, 55811957, 35695917, 265020, 265021, 265022, 60170615, 52644150, 264691, 264692, 33657023, 264693, 263966, 33657108, 27486281, 27486282, 27486284, 27486285, 35695763, 60431602, 18108370, 20281089, 264629, 18108374, 18108376, 55811578, 35696423, 35695855, 264634, 264635, 264636, 264655, 80431850, 264558, 264691
852	10147366 (1703, 1704)				

853	13032587 (1705, 1706)	Novel Protein sim. GBank gl 3402836 emb CAA76082  - (Y16136) 2-enoate reductase [Mooresella thermoacela]	reductase	264636 264566
854	80052438 (1707, 1708)			
855	79641130 (1709, 1710)			264692
856	11594236 (1711, 1712)		UNCLASSIFIED	264591
857	79210165 (1713, 1714)		UNCLASSIFIED	264630, 264634
858	80248910 (1715, 1716)			265008, 265009, 264601, 264602, 264603, 18108351
859	20296634 (1717, 1718)			264559
860	80041748 (1719, 1720)		UNCLASSIFIED	264489
861	65857045 (1721, 1722)		UNCLASSIFIED	33657023, 264630
862	80079467 (1723, 1724)			264600
863	80579931 (1725, 1726)	Novel Protein sim. GBank gl 2246532  (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]	UNCLASSIFIED	264488, 18108398, 35686286, 264259, 18108351, 264288, 265021
864	94839804 (1727, 1728)	Novel Protein sim. GBank gl 5889884 emb CAB52047.1  - (AL109732) hypothetical protein [Streptomyces coelicolor A3(2)]	UNCLASSIFIED	264259, 264112, 263974
865	80045310 (1729, 1730)			264635, 264600, 264636, 264591, 264602, 264693
866	80162031 (1731, 1732)	Novel Protein sim. GBank gl 4557676 ref NP_000341.1 pABCR - ATP binding cassette transporter	transport	264288, 264557, 264558
867	80062402 (1733, 1734)			264605
868	10075384 (1735, 1736)		UNCLASSIFIED	264909
869	80062408 (1737, 1738)			264605, 264687, 18108374
870	80249651 (1739, 1740)	Novel Protein sim. GBank gl 628660 p J537755 - Adenylyl-transferase - Escherichia coli	transferase	264601, 264636
871	20378295 (1741, 1742)	Novel Protein sim. GBank gl 1708180 sp Q10602 HEMK_MYCTU - HEMK PROTEIN HOMOLOG	UNCLASSIFIED	264603
872	95197114 (1743, 1744)	Novel Protein sim. GBank gl 1545959 emb CAA67763  - (X99384) paladin [Mus musculus]	UNCLASSIFIED	35686286, 22278998, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 35690052, 264509, 264905, 264908, 264907, 264908, 264909, 264510, 265006, 264511, 264512, 265007, 265008, 265009, 264910, 264591, 60433356, 264596, 52646317, 87168474, 265010, 264602, 264603, 265017, 265018, 264605, 18108351, 264764, 264766, 264768, 52644229, 264769, 21906765, 265021, 264534, 264691, 52645129, 264628, 264629, 35696423, 65274791, 264631, 264632, 264635, 264636, 264558, 264637, 264638, 264639, 80432113, 22278000, 22278002, 264564
873	20189728 (1745, 1746)	Novel Protein sim. GBank gl 4156104  (AE001569) - putative Outer membrane protein [Helicobacter pylori J95]	UNCLASSIFIED	264595

874	80077592 (1747, 1748)	Novel Protein sim. GBank gjl134319[sp]p07819[SCR8_BACSU - SUCROSE-6- PHOSPHATE HYDROLASE (SUCRASE)] (INVERTASE)		UNCLASSIFIED	264600
875	86608446 (1749, 1750)	Novel Protein sim. GBank gjl481000[pil]s37594 - mucin - human (fragment)			264259, 264448, 264288, 264557, 87168518
876	86465157 (1751, 1752)	Novel Protein sim. GBank gjl3128283 (AF010496) - iron(ii) diclrate transport ATP-binding protein [Rhodobacter capsulatus]	Contains protein domain (PF00005) - ABC transporter	transport	264907, 264601, 264602, 264605, 265020, 60431602
877	87802548 (1753, 1754)	Novel Protein sim. GBank gjl731074[sp]p40349[URB1 - SIDEROPHORE BIOSYNTHESIS REGULATORY PROTEIN URBS1]	Contains protein domain (PF00320) - GATA zinc finger	transcriptionfactor	22278998, 264908, 264369
878	80187289 (1755, 1756)	Novel Protein sim. GBank gjl1351614[sp]p09853[YAED_SCHPO - HYPOTHETICAL 181.5 KD PROTEIN C23D3.13C IN CHROMOSOME I]	ATPase-associated		264369, 264555
879	84328962 (1757, 1758)	Novel Protein sim. GBank gjl3875304[emb]CAA98434] - (Z74030) predicted using Genefinder; cDNA EST EMBL: C07608 comes from this gene; cDNA EST EMBL: C09023 comes from this gene; cDNA EST yk505e8.3 comes from this gene; cDNA EST yk489h9.3 comes from this gene; cDNA EST yk489h9.5 com...			56182575, 29331824, 264508, 264908, 265018, 18108351, 264448, 264883, 21906788, 21906788, 80170815, 33657023, 65274620, 33657109, 18108374, 35695855, 264563
880	8491135 (1759, 1760)	Novel Protein sim. GBank gjl137120[sp]p1214[UROT_MOUSE - TISSUE PLASMINOGEN ACTIVATOR PRECURSOR (TPA) (T- PLASMINOGEN ACTIVATOR)]	Contains protein domain (PF00051) - Kringle domain	cathepsin	264508
881	11290122 (1761, 1762)	Novel Protein sim. GBank gjl2632098[emb]CAA75667] - [Y15513] Prodos protein [Drosophila melanogaster]		UNCLASSIFIED	264508
882	11077011 (1763, 1764)	Novel Protein sim. GBank gjl1155068[emb]CAA64425] - [X94976] cell wall-plasma membrane linker protein [Brassica napus]		UNCLASSIFIED	264558
883	78582968 (1765, 1766)	Novel Protein sim. GBank gjl2078027[emb]CAB08487] - [Z95208] hypothetical protein Rv2372c [Mycobacterium tuberculosis]		UNCLASSIFIED	264688
884	13517921 (1767, 1768)	Novel Protein sim. GBank gjl2695834[emb]CAA15904] - [AL021006] sucA [Mycobacterium tuberculosis]		UNCLASSIFIED	264636
885	80052457 (1769, 1770)	Novel Protein sim. GBank gjl568935[db]BAA82981.1] - [AB028932] KIAA1029 protein [Homo sapiens]		UNCLASSIFIED	264605, 18108382
886	11685136 (1771, 1772)	Novel Protein sim. GBank gjl1881338[db]BAA19365] - [AB001488] PROBABLE INTEGRAL MEMBRANE PROTEIN, SIMILAR TO CHLORAMPHENICOL RESISTANCE PROTEIN OF STREPTOMYCES VENEZUELAE: [Bacillus subtilis]			264690
887	84315307 (1773, 1774)	Novel Protein sim. GBank gjl854065[emb]CAA58337] - [X83413] UBB [Human herpesvirus 6]		dehydrogenase	35696052, 264906, 264600, 264603, 35695917, 35695855, 264636
888	10083389 (1775, 1776)	Novel Protein sim. GBank gjl1881338[db]BAA19365] - [AB001488] PROBABLE INTEGRAL MEMBRANE PROTEIN, SIMILAR TO CHLORAMPHENICOL RESISTANCE PROTEIN OF STREPTOMYCES VENEZUELAE: [Bacillus subtilis]		UNCLASSIFIED	264908
889	20385917 (1777, 1778)	Novel Protein sim. GBank gjl1881338[db]BAA19365] - [AB001488] PROBABLE INTEGRAL MEMBRANE PROTEIN, SIMILAR TO CHLORAMPHENICOL RESISTANCE PROTEIN OF STREPTOMYCES VENEZUELAE: [Bacillus subtilis]			264603
890	19804337 (1779, 1780)	Novel Protein sim. GBank gjl1881338[db]BAA19365] - [AB001488] PROBABLE INTEGRAL MEMBRANE PROTEIN, SIMILAR TO CHLORAMPHENICOL RESISTANCE PROTEIN OF STREPTOMYCES VENEZUELAE: [Bacillus subtilis]			264629

881	13518878 (1781, 1782)	Novel Protein sim. GBank gi 4959398 gb AAD34331.1 AF11248 - (AF112481) RAD54B protein [Homo sapiens]		UNCLASSIFIED	264636
882	67634157 (1783, 1784)	Novel Protein sim. GBank gi 545526 bbs 143833 - LBP- 1betranscription factor binding to initiation site of HIV-1 (alternatively spliced) [human, Nematode cells, Peptide, 541 aa]		transcript factor	22278986, 22278999, 29331828, 35698052, 264808, 264809, 265009, 265011, 264802, 265019, 264766, 21906765, 21906766, 21906769, 265020, 285021, 56526488, 264689, 263967
883	78168037 (1785, 1786)	Novel Protein sim. GBank gi 2829888 sp P80608 CYSK_MAIZE - CYSTEINE SYNTHASE [O-ACETYL SERINE SULFHYDRYLASE] (O- ACETYL SERINE (THIO) L YASE) (CSASE)		synthase	
884	11102240 (1787, 1788)				263978
885	80239868 (1789, 1790)			UNCLASSIFIED	264508, 264600, 264555, 264559
886	79747803 (1791, 1792)				264632
887	94991923 (1793, 1794)			UNCLASSIFIED	264686, 29331828, 264511
888	67895109 (1795, 1796)		Contains protein domain (PF00047) - immunoglobulin domain		56182575, 60432289, 56182435, 60432229, 55811957, 22279000, 264486, 264601
889	11100463 (1797, 1798)	Novel Protein sim. GBank gi 1750127 (U66480) - YncC [Bacillus subtilis]		transport	264789, 264691, 264563
890	80499768 (1799, 1800)	Novel Protein sim. GBank gi 3122879 sp O07438 SYA_MYCTU - ALANYL-TRNA SYNTHETASE (ALANINE-TRNA LIGASE) (ALARS)	Contains protein domain (PF01411) - synthase (RNA synthetases class II (A))		264907, 264602, 264605, 264769, 35698017, 18108376, 264563
891	80502410 (1801, 1802)	Novel Protein sim. GBank gi 3122879 sp O07438 SYA_MYCTU - ALANYL-TRNA SYNTHETASE (ALANINE-TRNA LIGASE) (ALARS)			
892	80503301 (1803, 1804)	Novel Protein sim. GBank gi 335570 emb CAA200011 - (AL031124) 3-isopropylmalate dehydratase large subunit [Streptomyces coelicolor]	Contains protein domain (PF00330) - isomerase Aconitase family (aconitate hydratase)		264909, 265008, 264602, 264604, 264769, 264689, 264693
893	82060206 (1805, 1806)	Novel Protein sim. GBank gi 2960120 emb CAA18018.11 - (AL022121) glpK [Mycobacterium tuberculosis]	Contains protein domain (PF00370) - kinase FGG family of carbohydrate kinases		35698052, 264805, 264510, 264511, 264512, 264605, 264760, 18108351, 264762, 264687, 264768, 264769, 264688, 21906764, 35695917, 27486282, 35695855, 264634, 264636, 264486
894	20451078 (1807, 1808)	Novel Protein sim. GBank gi 28887 sp P40906 ARG_COCIM - ARGINASE	Contains protein domain (PF00491) - hydrolase Arginase family		264604
895	9398483 (1809, 1810)	Novel Protein sim. GBank gi 4567200 gb AAD23616.1 AC00716 - (AC007168) hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	264808
896	80052628 (1811, 1812)	Novel Protein sim. GBank gi 3738200 emb CAA212921 - (AL031855) putative vacuolar membrane protein [Schizosaccharomyces pombe]			264595, 264605
897	87813201 (1813, 1814)				
898	11754482 (1815, 1816)			UNCLASSIFIED	60432289, 264601, 264690
899	20727907 (1817, 1818)	Novel Protein sim. GBank gi 386894 gb BAA342961 - (AB015054) Alg2 [Rhizomucor pusillus]		UNCLASSIFIED	264638
900	16776205 (1819, 1820)	Novel Protein sim. GBank gi 4589726 gb BAA76883.11 - (AB003137) DnaJ homolog protein [Salix glauca]	Contains protein domain (PF00884) - eph DnaJ central domain (4 repeats)		264602

911	87454340 (1821, 1822)	Novel Protein sim. GBank gi 548774 sp P35685 RL7A_ORYSA - 60S RIBOSOMAL PROTEIN L7A	ribosomalprot	265010, 264604, 60432113
912	20448863 (1823, 1824)	Novel Protein sim. GBank gi 2314008 pp AAD07921.1  - (AE000597) CDP-diglyceride hydrolase (cdh) [Helicobacter pylori 28895]	hydrolase	264559
913	20469357 (1825, 1826)			
914	79183351 (1827, 1828)	Novel Protein sim. GBank gi 417657 sp Q03604 RIR1_CAEEL - PROBABLE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE LARGE CHAIN (RIBONUCLEOTIDE REDUCTASE)	UNCLASSIFIED	264604
915	87606703 (1829, 1830)	Novel Protein sim. GBank gi 568957 sp BAA83069.1  - (AB028040) KIAA1117 protein [Homo sapiens]	Contains protein domain (PF00317) - Ribonucleotide reductase	264636
916	79444081 (1831, 1832)	Novel Protein sim. GBank gi 4188110 emb CAA71790  - (Y10831) putative integrase [Ralstonia eutropha]	UNCLASSIFIED	18108398, 22278996, 66714117, 264906, 264591, 21906788, 265020, 55811576, 264638
917	20185985 (1833, 1834)	Novel Protein sim. GBank gi 1655699 emb CAA89032  - (Y07752) pherophorin-S [Volvox carter]	UNCLASSIFIED	264605
918	91226795 (1835, 1836)	Novel Protein sim. GBank gi 568966 emb CAB52005.1  - (AL109663) putative membrane protein [Streptomyces coelicolor A3(2)]	synthase	264258, 28331828, 264808, 265018, 264448, 265020, 264835, 83373044
919	80436785 (1837, 1838)	Novel Protein sim. GBank gi 568966 emb CAB52005.1  - (AL109663) putative membrane protein [Streptomyces coelicolor A3(2)]		265006, 264512, 264600, 264602, 264604, 264788, 18108370, 264583
920	79608095 (1839, 1840)	Novel Protein sim. GBank gi 1168448 sp Q05813 AMP1_STRLI - XAA-PRO AMINOPEPTIDASE I (X-PRO AMINOPEPTIDASE I) (AMINOPEPTIDASE P I) (APP) (PEPP I) (AMINOACYLPROLINE AMINOPEPTIDASE I)	peptidase	264508
921	19858534 (1841, 1842)	Novel Protein sim. GBank gi 3850084 emb CAA21911.1  - (AL033389) alcohol dehydrogenase [Schizosaccharomyces pombe]	UNCLASSIFIED	264600
922	78982505 (1843, 1844)	Novel Protein sim. GBank gi 287079 sp P29514 TB86_ARATH - TUBULIN BETA-8 CHAIN	tubulin	265019, 22279002
923	86695630 (1845, 1846)			264807, 265008, 285009, 285010, 18108351, 264688, 285021, 18108370, 18108374, 18108385
924	21431341 (1847, 1848)	Novel Protein sim. GBank		264510
925	20630332 (1849, 1850)	gi 2497688 sp Q06063 PAFA_MOUSE - PLATELET- ACTIVATING FACTOR ACETYLHYDROLASE PRECURSOR (PAF ACETYLHYDROLASE) (PAF 2- ACYLHYDROLASE) (LDL-ASSOCIATED PHOSPHOLIPASE A2) (LDL-PLA2) (2-ACETYL-1- ALKYLGLYCEROPHOSPHOCHOLINE ESTERASE) (1- ALKYL-2-ACETYLGLYCEROPHO...	esterase	264603
926	78397857 (1851, 1852)	Novel Protein sim. GBank gi 3882325 db BAA34522.1  - (AB018345) KIAA0802 protein [Homo sapiens]		55911957, 283972, 284639
927	37036201 (1853, 1854)		UNCLASSIFIED	264769



928	80070610 (1855, 1856)	Novel Protein sim. GBank gij155146 (M30318) - xanthine dehydrogenase (AA at 2538) [Calliphora vicina]		dehydrogenase	264605
929	20630336 (1857, 1858)	Novel Protein sim. GBank gij4115936[gj4AD03446.1] - (AF118223) No definition line found [Arabidopsis thaliana]		UNCLASSIFIED	284803
930	5496348 (1859, 1860)	Novel Protein sim. GBank gij4490609[emb]CAB36842.1] - (AJ133495) ribonucleotide reductase major subunit [Staphylococcus aureus]		reductase	264486
931	10245731 (1861, 1862)	Novel Protein sim. GBank gij5459396[emb]CAB50754.1] - (AL096839) putative integral membrane transport protein [Streptomyces coelicolor]		transport	284508, 284905, 284806, 264909, 264600, 264602, 264603, 264605, 264766, 264558, 18108387, 264486
932	80420813 (1863, 1864)	Novel Protein sim. GBank gij5689523[db]BAA83045.1] - (AB029016) KIAA1093 protein [Homo sapiens]		UNCLASSIFIED	284508, 264686, 264693, 27486261, 18108370, 65274791, 284636, 284559, 22279002
933	94326010 (1865, 1866)	Novel Protein sim. GBank gij1191111[sp]P12978[EBN2_EBV - EBNA-2 NUCLEAR PROTEIN]			264369, 22279002
934	80039105 (1867, 1868)	Novel Protein sim. GBank gij445686 (M32103) - ORF-27 [Staphylococcus aureus]		UNCLASSIFIED	264604, 264605, 264693, 18108370, 18108374
935	80063162 (1869, 1870)	Novel Protein sim. GBank gij1360666[db]CGHU1V - collagen alpha 1(V) chain precursor - human		UNCLASSIFIED	22278995, 22278996, 284602, 264687, 32833986, 18108387
936	80026632 (1871, 1872)	Novel Protein sim. GBank gij2226243[emb]CAA74531.1] - (Y14083) hypothetical protein [Bacillus subtilis]		synthase	263978
937	80250273 (1873, 1874)	Novel Protein sim. GBank gij3560166[emb]CAA20678] - (AL031525) ubiquitin carboxyl-terminal hydrolase [Schizosaccharomyces pombe]		UNCLASSIFIED	264602
938	80026633 (1875, 1876)	Novel Protein sim. GBank gij2494764[sp]Q50729[GUA, MYCTU - GMP SYNTHASE (GLUTAMINE-HYDROLYZING) (GLUTAMINE AMIDOTRANSFERASE) (GMP SYNTHETASE)]		ubiquitin	264600, 264908, 264907, 264908, 264909, 264511, 264910, 264592, 33657402, 264596, 264756, 284760, 264683, 264766, 264768, 264769, 33657023, 33657109, 264628, 264629, 264630, 264635, 264636, 264555, 264637, 284556, 264638, 284639, 83373044, 18108385, 264565, 18108391
939	11071694 (1877, 1878)	Novel Protein sim. GBank gij4033729 (AF038595) - apolipoprotein N-acyltransferase [Pseudomonas aeruginosa]			284593
940	94144252 (1878, 1880)	Novel Protein sim. GBank gij2494764[sp]Q50729[GUA, MYCTU - GMP SYNTHASE (GLUTAMINE-HYDROLYZING) (GLUTAMINE AMIDOTRANSFERASE) (GMP SYNTHETASE)]		UNCLASSIFIED	264760
941	11398414 (1881, 1882)	Novel Protein sim. GBank gij732353[sp]P39606[YWCH, BACSU - HYPOTHETICAL 35.6 KD PROTEIN IN OOXD-VPR INTERGENIC REGION]		UNCLASSIFIED	264600, 264687, 264689, 264563
942	19484122 (1883, 1884)	Novel Protein sim. GBank gij2494764[sp]Q50729[GUA, MYCTU - GMP SYNTHASE (GLUTAMINE-HYDROLYZING) (GLUTAMINE AMIDOTRANSFERASE) (GMP SYNTHETASE)]		synthase	264511, 264603
943	80060258 (1885, 1886)	Novel Protein sim. GBank gij732353[sp]P39606[YWCH, BACSU - HYPOTHETICAL 35.6 KD PROTEIN IN OOXD-VPR INTERGENIC REGION]		UNCLASSIFIED	264906, 264604, 264605, 265020, 18108387
944	80216006 (1887, 1888)	Novel Protein sim. GBank gij732353[sp]P39606[YWCH, BACSU - HYPOTHETICAL 35.6 KD PROTEIN IN OOXD-VPR INTERGENIC REGION]		UNCLASSIFIED	265017
945	80052477 (1889, 1890)	Novel Protein sim. GBank gij732353[sp]P39606[YWCH, BACSU - HYPOTHETICAL 35.6 KD PROTEIN IN OOXD-VPR INTERGENIC REGION]		UNCLASSIFIED	265017
946	78248402 (1891, 1892)	Novel Protein sim. GBank gij732353[sp]P39606[YWCH, BACSU - HYPOTHETICAL 35.6 KD PROTEIN IN OOXD-VPR INTERGENIC REGION]		UNCLASSIFIED	265017

947	81802598 (1893, 1894)	Novel Protein sim. GBank gjl2896770[emb]CAA172471 - (AL021899) hypothetical protein Rv2033c [Mycobacterium tuberculosis]	Contains protein domain (PF00459) - inositol monophosphatase family	phosphatase	18108394, 22278956, 264907, 264909, 285008, 265009, 284910, 264758, 264600, 264602, 265018, 264605, 284789, 264689, 264693
948	88165538 (1895, 1896)	Novel Protein sim. GBank gjl2827284 (AF041037) - novel antagonist of FGF signaling [Homo sapiens]		lgf	18108398, 56182575, 22278997, 22278999, 60432049, 29331822, 29331826, 264907, 56182435, 55811386, 285011, 264600, 265017, 265018, 265019, 18108351, 265020, 265021, 265022, 27486265, 263972, 55811576, 264638, 80170394, 264566
949	88081786 (1897, 1898)	Novel Protein sim. GBank gjl4507985[ref]NP_003427.1[pZNF1 - zinc finger protein 135 (clone pHZ-17)]	Contains protein domain (PF00095) - Zinc finger, C2H2 type	transcription factor	29331825, 21905764, 27486261, 21905766, 52844296, 33657349, 87188518, 56994079, 265020, 265021, 87188599, 52644150, 264637
950	79485872 (1899, 1900)	Novel Protein sim. GBank gjl1079481[pil]S43865 - cytochrome b, type II - pitoroc (fragment)	Contains protein domain (PF00038) - intermediate filament proteins	struct	284883, 18108361
951	20451411 (1901, 1902)	Novel Protein sim. GBank gjl5420387[emb]CAB46679.11 - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264604
952	79566954 (1903, 1904)	Novel Protein sim. GBank gjl5305702[gb]AAD41779.1[pF12686 - (AF126867) calpain-like protease [Mus musculus]		calhepsin	264910, 264691
953	10196003 (1905, 1906)	Novel Protein sim. GBank gjl2495642[sp]Q47142[YFHS_ECOLI - HYPOTHETICAL 41.8 KD PROTEIN IN CSIE-GLYA INTERGENIC REGION]		transport	284510
954	9893326 (1907, 1908)	Novel Protein sim. GBank gjl2360965 (AF016253) - D-amino acid dehydrogenase [Klebsiella aerogenes]		dehydrogenase	284508
955	95313410 (1909, 1910)	Novel Protein sim. GBank gjl5454064[ref]NP_006319.1[pSIP] - SYT interacting protein	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	dna_rna_bind	56994075, 264509, 264905, 264906, 264907, 264908, 264909, 284510, 264758, 264759, 265010, 264601, 284760, 18108351, 284782, 284783, 284764, 284788, 264686, 284787, 284887, 284788, 284789, 264689, 284628, 264629, 264630, 264631, 264632, 284634, 284635, 284636, 264637, 284638, 56182323, 264639, 18108388, 284563, 284564
956	80084224 (1811, 1912)	Novel Protein sim. GBank gjl2052129[emb]CAB081551 - (Z94752) rimJ [Mycobacterium tuberculosis]			284605
957	80056208 (1913, 1914)			UNCLASSIFIED	264603, 18108362
958	80036448 (1915, 1916)	Novel Protein sim. GBank gjl1709787[sp]Q00451[PRF1_LYCES - 36.4 KD PROLINE-RICH PROTEIN]		UNCLASSIFIED	264908, 264910, 264762, 263978, 264637
959	80026647 (1917, 1918)	Novel Protein sim. GBank gjl2131050[emb]CAB092601 - (Z85944) opca [Mycobacterium tuberculosis]		UNCLASSIFIED	264602, 264692
960	37815408 (1919, 1920)	Novel Protein sim. GBank gjl2129478[pil]S51939 - chitinase (EC 3.2.1.14) precursor - beet	/	UNCLASSIFIED	264259
961	20567283 (1921, 1922)				263978
962	11399318 (1923, 1924)				284593

963	80590374 (1925, 1926)				UNCLASSIFIED	264510, 264288, 264555, 264556, 264559, 264486
964	79832019 (1927, 1928)	Novel Protein sim. GBank gi458622[dbjBAA76833.1] - (AB023206) KIAA0989 protein [Homo sapiens]			UNCLASSIFIED	264112, 264010, 264689
965	91229485 (1929, 1930)	Novel Protein sim. GBank gi5420387[emb CAB46679.1] - (AJ243450) proteophosphoglycan [Leishmania major]			UNCLASSIFIED	264488, 265017, 264448, 264634, 264558, 83373044
966	95292815 (1931, 1932)				UNCLASSIFIED	264908, 264592, 264596, 264604, 264768, 21905764, 264692, 264693, 264629, 264636, 264638
967	79258708 (1933, 1934)	Novel Protein sim. GBank gi1731207[sp Q11156 RCX3_MYCTU - SENSORY TRANSDUCTION PROTEIN REGX3	Contains protein domain (PF00072) - Response regulator receiver domain		phosphatase	264750
968	79560268 (1935, 1936)	Novel Protein sim. GBank gi1266183[emb CAA75187.1] - (Y14864) putative transport protein [Methylophilus methylotrophus]			transport	264693
969	79918470 (1937, 1938)	Novel Protein sim. GBank gi1541887[emb CAB46422.1] - (AL098747) hypothetical protein [Homo sapiens]	Contains protein domain (PF00098) - Zinc finger, C2H2 type		dna_ma_bind	35698286, 264665, 264686, 35695917, 264692, 18108374, 264635
970	95085947 (1939, 1940)				UNCLASSIFIED	22278996, 22278998, 22278999, 28147820, 264828, 265006, 265007, 265008, 265009, 18108348, 33109934, 265010, 265011, 18108351, 264288, 21906767, 21906768, 18108370, 18108374, 18108377, 264630, 264635, 18108380, 83373044, 18108387, 18108388
971	78918770 (1941, 1942)				UNCLASSIFIED	265007, 265020, 22279002
972	20710704 (1943, 1944)					264557
973	20370183 (1945, 1946)	Novel Protein sim. GBank gi1723119[sp P53990 Y174_HUMAN - HYPOTHETICAL PROTEIN KIAA0174				264604
974	80057103 (1947, 1948)				UNCLASSIFIED	264555
975	10186018 (1948, 1950)				UNCLASSIFIED	264510
976	80205742 (1951, 1952)	Novel Protein sim. GBank gi3861459[emb CAA92988.1] - (Z68753) predicted using GeneFinder: Similarity to Yeast hypothetical protein YIK9 (SW:YIK9_YEAST); cDNA EST EMBL:D27680 comes from this gene; cDNA EST EMBL:D27679 comes from this gene; cDNA EST EMBL:D64477 comes from this ge...			UNCLASSIFIED	264508, 264906, 264758, 264632, 264639, 264583
977	10355349 (1953, 1954)	Novel Protein sim. GBank gi1549458[sp Q05335 XY33_PSEPU - XYLDLEGF OPERON TRANSCRIPTIONAL ACTIVATOR 3			UNCLASSIFIED	264906
978	80025927 (1955, 1956)				UNCLASSIFIED	264800, 264602, 264603, 264604
979	80447820 (1957, 1958)	Novel Protein sim. GBank gi3171904[emb CAA75869.1] - (Y15908) DIA-12C protein [Homo sapiens]			UNCLASSIFIED	264767, 264768, 265006, 265007, 264906
980	80025928 (1959, 1960)				UNCLASSIFIED	264600, 264602, 264605
981	80098550 (1961, 1962)	Novel Protein sim. GBank gi3599940 (AF017388) - facogential dysplasia protein 2 [Mus musculus]			UNCLASSIFIED	264692, 264555, 264556, 264557, 264559

982	80185670 (1983, 1984)	Novel Protein sim. GBank gll2950220[emb]CAA71575] - (Y10545) fused-ccdB [Escherichia coli]		UNCLASSIFIED	264404
983	90995041 (1985, 1986)	Novel Protein sim. GBank gll476389[pr]B43402 - myosin heavy chain-B, neuronal - chicken		struct	65274572, 56182575, 264908, 264909, 265007, 265008, 264758, 265010, 59811150, 33657023, 264834, 264557, 264558
984	20468876 (1987, 1988)	Novel Protein sim. GBank gll3451504[emb]CAA07660.1] - (AJ007747) hypothetical protein BbLPS1.21 [Bordetella bronchiseptica]	Contains protein domain (PF00534) - Glycosyl transferases group 1	UNCLASSIFIED	264605
985	65461368 (1989, 1990)			transferase	56182435, 264600
986	87102888 (1971, 1972)			UNCLASSIFIED	264106, 264110, 265020, 60170615
987	78867231 (1973, 1974)			UNCLASSIFIED	264909
988	18658681 (1975, 1976)			UNCLASSIFIED	264600
989	88095329 (1977, 1978)			UNCLASSIFIED	264508, 265017, 264534, 264564
990	88057746 (1979, 1980)	Novel Protein sim. GBank gll5725506[gb]AAD46080.1]AF06015 - (AF060152) METH1 protein [Homo sapiens]	Contains protein domain (PF01421) - Reprolysin (M128) family zinc metalloprotease	UNCLASSIFIED	264258, 264908, 265009, 264910, 264596, 264368, 264288, 264768, 264628, 264635, 264588
991	10106140 (1981, 1982)	Novel Protein sim. GBank gll2105049[emb]CAB08835] - (Z85436) hypothetical protein RV3845 [Mycobacterium tuberculosis]	Contains protein domain (PF00211) - Adenylate and Guanylate cyclase catalytic domain	UNCLASSIFIED	264909
992	70845694 (1983, 1984)			UNCLASSIFIED	264508, 264593
993	10814053 (1985, 1986)	Novel Protein sim. GBank gll3329297 (AE001355) - Ribonucleoside Reductase, Large Chain [Chlamydia trachomatis]		reductase	264907
994	11090590 (1987, 1988)				264602
995	94321911 (1988, 1990)	Novel Protein sim. GBank gll5106572[gb]AAD39780.1]AF14394 - transcriptional activator SRCAP [Homo sapiens]	Contains protein domain (PF00176) - SNF2 and others N-terminal domain	helicase	18108398, 55274572, 22278996, 264490, 60432049, 29331827, 29148498, 264508, 264905, 264907, 264908, 58182435, 265008, 264591, 264592, 60432228, 60431735, 33657402, 264595, 264768, 21908754, 265010, 265017, 265018, 264605, 264760, 284448, 264763, 264768, 21908765, 21908766, 21908769, 55811957, 264692, 264693, 264629, 35696423, 55811576, 35695855, 264636, 264555, 264556, 264558, 83373044, 22278002, 284593
996	91013745 (1991, 1992)	Novel Protein sim. GBank gll2911719 (AC004227) - KIA001LB [Homo sapiens]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF)	phosphatase	83373044, 29331824, 263978, 55811957, 56526488, 87188518, 264910, 264908, 264595, 264586, 264693, 264766
997	80503347 (1993, 1994)	Novel Protein sim. GBank gll2649101 (AE001001) - ABC transporter, ATP-binding protein [Archaeoglobus fulgidus]	Contains protein domain (PF00005) - ABC transporter	transport	35696286, 22278997, 22278999, 264508, 264905, 264908, 265010, 264800, 264802, 264605, 264688, 264769, 265021, 264555, 18108391
998	11397390 (1995, 1996)	Novel Protein sim. GBank gll123530[sp]P04929[HRPX_PLALO - HISTIDINE-RICH GLYCOPROTEIN PRECURSOR			264595

999	11768047 (1997, 1998)	Novel Protein sim. GBank gij2506897 sp P46490 YFCA_HAEIN - HYPOTHETICAL PROTEIN H10198		UNCLASSIFIED	264682	
1000	20727844 (1999, 2000)			UNCLASSIFIED	264602	
1001	86673131 (2001, 2002)	Novel Protein sim. GBank gij224689 dbj BAA20833  - (AB002377) KIAA0379 [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	kinase	60432049, 264907, 264909, 264511, 264603, 264683, 264684, 264687, 264689, 26148627, 21906769, 264692, 18108385, 22279000 265009, 264389, 265020	
1002	80189603 (2003, 2004)	Novel Protein sim. GBank gij586121 sp P37709 TRHY_RABIT - TRICHOHYALIN		struct		
1003	17833491 (2003, 2006)				265019	
1004	16314987 (2007, 2008)	Novel Protein sim. GBank gij854065 emb CAA58337  - (X83413) U88 [Human herpesvirus 6]			264635	
1005	79617144 (2008, 2010)	Novel Protein sim. GBank gij114073 sp P07672 APT_ECOLI - ADENINE PHOSPHORIBOSYLTRANSFERASE (APRT)		UNCLASSIFIED	264508	
1006	37815429 (2011, 2012)			UNCLASSIFIED	264259	
1007	78620871 (2013, 2014)	Novel Protein sim. GBank gij406297 dbj BAA36210.1  - (AB017138) epsilon subunit of malonate decarboxylase [Pseudomonas putida]		synthase	264905	
1008	86094444 (2015, 2016)	Novel Protein sim. GBank gij2808807 emb CAA04607.1  - (AJ001206) putative trehalose synthase [Streptomyces coelicolor]		synthase	265007, 264602, 264605, 264760, 264636	
1009	57451288 (2017, 2018)	Novel Protein sim. GBank gij3639077 (AF090113) - AMPA receptor binding protein [Rattus norvegicus]	Contains protein domain (PF00595) - POZ domain (Also known as DHR or GLGF).	kinase	264102, 264288	
1010	94672537 (2019, 2020)	Novel Protein sim. GBank gij3746332 (AF016307) - possible NADH-dependent oxidase, may function as a demethylase [Sinorhizobium meliloti]		dehydrogenase	264592	
1011	85546916 (2021, 2022)	Novel Protein sim. GBank gij2342647 dbj AA885591.1  - (U90653) DRHC-domain-containing cysteine-rich protein [Homo sapiens]		UNCLASSIFIED	35696052, 264905, 284764, 284768, 35695917, 264628	
1012	85284456 (2023, 2024)	Novel Protein sim. GBank gij341341 emb CAA20272  - (AL031231) guanosine pentaphosphate synthase/ polyribonucleotide nucleotidyltransferase [Streptomyces coelicolor]	Contains protein domain (PF00013) - KH domain	phosphorylase	35696052, 264905, 264600, 264601, 264602, 264605, 264762, 264766, 284768, 284689	
1013	86095772 (2025, 2026)				264591, 21906768	
1014	86808828 (2027, 2028)			UNCLASSIFIED	29331824, 285019, 265020	

1015	95418679 (2029, 2030)	Novel Protein sim. GBank gll4159995 (AF063095) - SELL [Mus musculus]	Contains protein domain (PF00040) - Struct Fibrinectin type II domain		22278994, 22278995, 56994075, 22278996, 22278999, 264259, 29331825, 29331828, 264907, 56182435, 284510, 264591, 264593, 60433396, 264594, 55812038, 264758, 21906754, 33657084, 265010, 264600, 265017, 265018, 265019, 18108351, 21906765, 21906766, 21906767, 21906768, 55811857, 265022, 33657023, 65274620, 33657182, 32833988, 18108370, 18108377, 55811576, 33690423, 264630, 22279000, 264565
1016	79559694 (2031, 2032)	Novel Protein sim. GBank gll2506969sp141407IACPD_ECOLI - ACYL CARRIER PROTEIN PHOSPHODIESTERASE (ACP PHOSPHODIESTERASE)	esterase		264686, 264693
1017	11069213 (2033, 2034)	Novel Protein sim. GBank gll5103943dbjBAA79259.1j - protein oppA [Aeropyrum pernix]	transport		264600
1018	80072430 (2035, 2036)	Novel Protein sim. GBank gll4493973lmbjCAB39032.1j - (AL034559) predicted using hexExon; MAL3P7.14 (PFC0925w). Hypothetical protein, len: 489 aa [Plasmodium falciparum]	Contains protein domain (PF00496) - Bacterial extracellular solute-binding proteins, family 5		22278995, 29148627, 264563
1019	11703607 (2037, 2038)		UNCLASSIFIED		264688
1020	80234432 (2039, 2040)				264508, 264509, 264512, 264800, 264762, 264769, 264689, 18108370, 264636, 264638, 264486
1021	37036243 (2041, 2042)	Novel Protein sim. GBank gll4633607lgbjAAD28859.1jAF12779 - (AF127795) trehalose biosynthetic enzyme TreY [Rhizobium leguminosarum bv. viciae]	synthase		264769
1022	80502627 (2043, 2044)	Novel Protein sim. GBank gll1781230lmbjCAB052771 - (Z83867) hypothetical protein Rv3137 [Mycobacterium tuberculosis]	phosphatase		33696052, 264508, 265008, 265009, 264769, 18108387, 264563
1023	11399341 (2045, 2046)	Novel Protein sim. GBank gll3777495 (U92083) - calcium transporting ATPase [Pichia angusta]	Contains protein domain (PF00122) - E1-E2 ATPase		264593
1024	80057129 (2047, 2048)		UNCLASSIFIED		52846842, 33657402, 33657023, 18108379, 55811576, 264631, 264556, 264557, 264559, 18108386, 264566
1025	79644200 (2049, 2050)	Novel Protein sim. GBank gll3483045lmbjCAA205561 - (AL031371) putative transport system permease protein [Streptomyces coelicolor]	transport		264693
1026	80025946 (2051, 2052)	Novel Protein sim. GBank gll174922lspjQ03322UVRD_HAEIN - DNA HELICASE II	helicase		264602
1027	17899234 (2053, 2054)	Novel Protein sim. GBank gll4757728lrefjNP_004886.1pAGTA - angiotensin vasopressin receptor AII/AVP-like	UNCLASSIFIED		265017

1028	20257928 (2055, 2056)	Novel Protein sim. GBank gij2791409[emb CAA16003] - (AL021184) aen [Mycobacterium tuberculosis]	Contains protein domain (PF00330) - Aconitase family (aconitase hydratase)	UNCLASSIFIED	284600
1029	94655090 (2057, 2058)			UNCLASSIFIED	284595
1030	86095343 (2059, 2060)			UNCLASSIFIED	284907, 284908, 284510, 284512, 285008, 285010, 285011, 284600, 284602, 284603, 284605, 284769, 18108372, 18108374
1031	95289117 (2061, 2082)			UNCLASSIFIED	284905, 284906, 284909, 284595, 284692, 284630, 284634, 284638
1032	94613275 (2063, 2064)	Novel Protein sim. GBank gij4503895[ref NP_000145.1]pGALK - galactokinase 1		UNCLASSIFIED	284688
1033	86464818 (2065, 2066)	Novel Protein sim. GBank gij2982990 (AE000682) - hypothetical protein [Aquifex aeolicus]		UNCLASSIFIED	35896052, 284908, 284510, 18108354, 284687, 284769, 284689, 60431602, 18108385, 284486
1034	79245937 (2067, 2068)	Novel Protein sim. GBank gij405895 (U00007) - methionyl-tRNA synthetase [Escherichia coli]		UNCLASSIFIED	284908
1035	79956355 (2069, 2070)			UNCLASSIFIED	284682
1036	85804988 (2071, 2072)			UNCLASSIFIED	284905, 66712502, 284808, 284768
1037	87896058 (2073, 2074)			UNCLASSIFIED	29331824, 284909, 60433438, 285019
1038	20481015 (2075, 2076)	Novel Protein sim. GBank gij790819 (L38881) - polycystic kidney disease-associated protein [Homo sapiens]	Contains protein domain (PF01477) - PLAT/LH2 domain		284604, 284634
1039	87260021 (2077, 2078)	Novel Protein sim. GBank gij2605967 (AF030027) - 24 [Equine herpesvirus 4]		UNCLASSIFIED	264082, 264093, 264094, 264683, 264689, 283967
1040	80026640 (2079, 2080)	Novel Protein sim. GBank gij2352095 (U97022) - DNA topoisomerase I [Fervidobacterium islandicum]	Contains protein domain (PF01131) - Prokaryotic DNA topoisomerase	isomerase	284595
1041	10156682 (2081, 2082)	Novel Protein sim. GBank gij3256535[dbj BA029218.1] - (AP000001) 301aa long hypothetical 2-phosphoglycerate kinase [Pyrococcus horikoshii]	kinase		284907
1042	11084375 (2083, 2084)	Novel Protein sim. GBank gij2058289[emb CAA68953] - (X98309) ARI protein [Drosophila melanogaster]			284605
1043	80037138 (2085, 2086)	Novel Protein sim. GBank gij1870167[emb CAA70125] - (Y08921) msK [Streptomyces reliculi]	Contains protein domain (PF00005) - ABC transporter	transport	284565, 264567
1044	80025952 (2087, 2088)	Novel Protein sim. GBank gij5689890[emb CAE52053.1] - (AL109732) hypothetical protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	285006, 284802, 285017
1045	52415482 (2089, 2090)	Novel Protein sim. GBank gij854065[emb CAA58337] - (X83413) U88 [Human herpesvirus 6]	helicase		284686
1046	11754662 (2091, 2092)	Novel Protein sim. GBank gij4210471[dbj BA074535.1] - (AB018033) orfSA [Pseudomonas sp.]		UNCLASSIFIED	284769
1047	37036258 (2093, 2094)	Novel Protein sim. GBank gij341341[emb CAA20279] - (AL031232) hypothetical protein SC10H5.07 [Streptomyces coelicolor]	Contains protein domain (PF00220) - Neurohypophysial hormones, N-terminal Domain	UNCLASSIFIED	284687
1048	81755108 (2097, 2098)	Novel Protein sim. GBank gij5051636[gb AAD3828.1]AF07372 - (AF07372) EH domain-binding mitotic phosphoprotein [Homo sapiens]		UNCLASSIFIED	264905, 284634
1050	79471521 (2099, 2100)			UNCLASSIFIED	284686

1051	80475471 (2101, 2102)				UNCLASSIFIED	18108374, 264769, 285010, 285011, 284801, 265009, 264604, 264605, 264636, 18108351, 264692, 264508, 264762, 264687, 264486
1052	82442862 (2103, 2104)	Novel Protein sim. GBank gi 3123275 sp P35136 SERA_BACSU - D-3- PHOSPHOGLYCERATE DEHYDROGENASE (PGDH)		Contains protein domain (PF00389) - D-isomer specific 2-hydroxyacid dehydrogenases		
1053	94851640 (2105, 2106)	Novel Protein sim. GBank gi 5441319 emb CAB48717.1  - (AL034396) dJ158812.1 (zinc finger, X-linked, duplicated A) [Homo sapiens]				264888, 18108374, 28331824, 33373044, 21908754, 52645156, 58182435, 264689, 29331827, 27486261, 35698052, 21908765, 35698423, 21908768, 58182575, 21908769, 55811957, 87168518, 35686286, 22278997, 285020, 285011, 285021, 285022, 285007, 285018, 22279000, 22279002, 264482, 264906, 52644150, 264909, 264288, 29331822, 52645080, 264766
1054	79580225 (2107, 2108)	Novel Protein sim. GBank gi 5052508 gb AA038584.1 AF145609			UNCLASSIFIED	264686
1055	80594138 (2109, 2110)	BcDNA_QH-02833 [Drosophila melanogaster]		Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	284907, 264602, 264681, 264288, 21906768, 33857109, 55810764, 35685855, 264631
1056	17882319 (2111, 2112)	Novel Protein sim. GBank gi 3021676 db BA25358  - (D86033) RNA polymerase sigma-70 factor [Pseudomonas fluorescens]			mapolymerase	264906
1057	85667216 (2113, 2114)	Novel Protein sim. GBank gi 1226281 (U50308) - No definition line found [Caenorhabditis elegans]			UNCLASSIFIED	264682
1058	80376376 (2115, 2116)	Novel Protein sim. GBank				264764
1059	84662754 (2117, 2118)	gi 170016 sp P46808 GREM_MYCLE - TRANSCRIPTION ELONGATION FACTOR GREM (TRANSCRIPT CLEAVAGE FACTOR GREM)			transcript factor	35696052, 35695555, 265009, 264636
1060	78481169 (2119, 2120)	Novel Protein sim. GBank gi 2499087 sp Q09332 UGGG, DROME - UDP- GLUCOSE GLYCOPROTEIN GLUCOSYLTRANSFERASE PRECURSOR (DUGT)			glycoprotein	29146499, 264681, 264683, 264687
1061	11034025 (2121, 2122)	Novel Protein sim. GBank gi 90254 pi j128334 - protein- tyrosine-phosphatase (EC 3.1.3.48) Ly-5 precursor (B-cell variant) - mouse			phosphatase	264634
1062	35587937 (2123, 2124)	Novel Protein sim. GBank gi 3334200 sp Q49954 GCSP_SOLTU - GLYCINE DEHYDROGENASE (DECARBOXYLATING) PRECURSOR (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-PROTEIN)			dehydrogenase	284593
1063	8490481 (2125, 2126)	Novel Protein sim. GBank gi 2499986 sp Q41228 PSE1_NICSY - PHOTOSYSTEM I REACTION CENTRE SUBUNIT IV A PRECURSOR (PSI-E A)				264508
1064	78891783 (2127, 2128)	Novel Protein sim. GBank gi 82654 pi j1A0086 - 10K zein precursor - maize				265007, 265008, 18108351, 18108385



1065	80021208 (2128, 2130)	Novel Protein sim. GBank gi 2120988 pir S70682 - glycosyltransferase homolog - Bordetella pertussis		transferase	264600, 264602, 264689
1066	17896879 (2131, 2132)	Novel Protein sim. GBank gi 2506362 sp P15042 DNLL_ECOLI - DNA LIGASE (POLYDEOXYRIBONUCLEOTIDE SYNTHASE (NAD+))		synthase	265009
1067	10132178 (2133, 2134)				264909
1068	82062057 (2135, 2136)	Novel Protein sim. GBank gi 4007668 emb CAA22355  - (AL034443) putative oxidoreductase (Streptomyces coelicolor)	Contains protein domain (PF00248) - Aldol/keto reductase family	reductase	264688, 18108362, 264558, 264600, 264760
1069	83002954 (2137, 2138)	Novel Protein sim. GBank gi 4589484 gb BAA76770.1  - (AB023143) KIAA0928 protein [Homo sapiens]		UNCLASSIFIED	264604
1070	82101992 (2139, 2140)	Novel Protein sim. GBank gi 120304 sp P15932 FLGK_SALTY - FLAGELLAR HOOK-ASSOCIATED PROTEIN 1 (HAP1)		UNCLASSIFIED	264604, 264760
1071	20710589 (2141, 2142)	Novel Protein sim. GBank gi 1750387 U81281 - glutamate synthase large subunit [Pseudomonas aeruginosa]		synthase	264602
1072	82356540 (2143, 2144)				264687, 264688, 21906784, 35695052, 35695917, 35695895, 264600, 264601, 264602, 265009, 264605, 264508, 264905, 264690, 264906, 264762, 264628, 264768
1073	79814400 (2145, 2146)	Novel Protein sim. GBank gi 477532 pir A49175 - Motch B protein - mouse (fragment)	Contains protein domain (PF00008) - EGF-like domain	synthase	264909
1074	80105992 (2147, 2148)	Novel Protein sim. GBank gi 3893109 emb CAA76940  - (Y17920) CALO protein [Drosophila melanogaster]		UNCLASSIFIED	56894075, 22278988, 264594, 264757, 264596, 265018, 265019, 264681, 264389, 264688, 265020, 18108364, 18108374
1075	81850293 (2149, 2150)			ATPase-associated	264769
1076	80477264 (2151, 2152)	Novel Protein sim. GBank gi 1176203 sp P46442 YHGM_ECOLI - HYPOTHETICAL 43.1 KD PROTEIN IN RPLM-HHOA INTERGENIC REGION (F375)			
1077	79831334 (2153, 2154)	Novel Protein sim. GBank gi 4033487 sp Q44472 TUD4_AGRVI - PUTATIVE HYDROXYPYRUVATE REDUCTASE		kinase	264905
1078	20288674 (2155, 2156)				
1079	80494518 (2157, 2158)	Novel Protein sim. GBank gi 3413828 emb CAA20296  - (AL031260) hypothetical protein SC9A10.09 [Streptomyces coelicolor]		UNCLASSIFIED	264600
1080	11767183 (2159, 2160)			UNCLASSIFIED	18108394, 264769, 264634, 264636
1081	94747080 (2161, 2162)			UNCLASSIFIED	264684
1082	81490655 (2163, 2164)			UNCLASSIFIED	83373044, 265019, 22278002, 264482, 18108351, 264682, 264906, 264693, 264487, 264758, 264788, 264789, 21806787, 264511, 264810, 264634, 264635, 264905, 264636, 264906, 264637, 264907, 264908, 264704, 264638, 20281099, 264766, 264595

1083	87446717 (2165, 2166)	Novel Protein sim. GBank gjl1722945[sp]Q10523[Y01N_MYCTU - HYPOTHETICAL 44.6 KD PROTEIN CY427.23]		UNCLASSIFIED	60422479, 264905, 264906, 264510, 60432229, 264759, 87168474, 264603, 264766, 264689, 18108384, 18108376, 35695855, 264636 264769
1084	37789308 (2187, 2188)	Novel Protein sim. GBank gjl418384[sp]P32057[MCAL_ECOLI - PUTATIVE COLANIC ACID BIOSYNTHESIS GLYCOSYL TRANSFERASE WCAL]		UNCLASSIFIED	
1085	86475368 (2169, 2170)	Novel Protein sim. GBank gjl1899190 (U80204) - heat shock protein 60 [Tsukamurella lyosinsidivens]	Contains protein domain (PF00118) - eph TCP-1/cpn60 chaperonin family		60432229, 264687
1086	79808269 (2171, 2172)	Novel Protein sim. GBank gjl112856[sp]P46176[RL14_ACYKS - 50S RIBOSOMAL PROTEIN L14]	Contains protein domain (PF00238) - ribosomalprot Ribosomal protein L14		264486
1087	79603979 (2173, 2174)	Novel Protein sim. GBank gjl4160188[emb]CAA15431 - (AL008583) dJ327J16.3 (novel CHROMOBOX family protein)	Contains protein domain (PF00385) - 'chromo' (CHRromatin Organization Modifier) domain		29331827, 264693
1088	79854983 (2175, 2176)	Novel Protein sim. GBank gjl2983155 (AE000693) - phosphoglucosyltransferase [Aquifex aerophilus]		UNCLASSIFIED	264905, 264601, 18108387
1089	80216600 (2177, 2178)	Novel Protein sim. GBank gjl4981768[sp]AD36280.1(AE00177 - (AE001776) NADH dehydrogenase, 30 kDa subunit, putative [Thermotoga maritima]	Contains protein domain (PF00329) - Respiratory-chain NADH dehydrogenase, 30 Kd subunit	UNCLASSIFIED	264488, 264511, 265011, 264682, 264768, 264689, 21906764, 35695917, 265020, 32833986, 18108370, 35695855
1090	11083825 (2179, 2180)	Novel Protein sim. GBank gjl4007680[emb]CAA223661 - (AL034443) putative oxidoreductase [Streptomyces coelicolor]			264604
1091	12917471 (2181, 2182)	Novel Protein sim. GBank gjl2495582[sp]P77239[YLCD_ECOLI - HYPOTHETICAL 44.3 KD PROTEIN IN NFRB-PREP INTERGENIC REGION PRECURSOR]		UNCLASSIFIED	264637
1092	80252286 (2183, 2184)	Novel Protein sim. GBank gjl2960098[emb]CAA17996.1 - (AL022121) ntl [Mycobacterium tuberculosis]	Contains protein domain (PF00730) - Endonuclease III		264556 264769, 35695917, 35695855, 264600, 264602, 264803, 264605, 18108351
1093	80496304 (2185, 2186)	Novel Protein sim. GBank gjl1001642[db]BAA103731 - (D84002) dGTP triphosphohydrolase [Synecocystis sp.]		UNCLASSIFIED	264686
1094	10880972 (2187, 2188)	Novel Protein sim. GBank gjl4585587[emb]CAB40855.1 - (AL049628) putative adenine glycosylase [Streptomyces coelicolor]	Contains protein domain (PF00455) - Bacterial regulatory proteins, deoR family		264906, 265007, 264595, 264600, 264602, 264603, 264604, 264605, 264762, 264768, 264769, 264636, 264558, 18108387, 60432113, 264482, 264485
1095	87457250 (2189, 2190)	Novel Protein sim. GBank gjl115001[sp]P19206[IOB_BACSH - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)]			264600, 264602, 264603, 264604, 264605, 35695917, 264692, 264631
1096	80025977 (2191, 2192)	Novel Protein sim. GBank gjl115001[sp]P19206[IOB_BACSH - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)]			
1097	79239560 (2193, 2194)	Novel Protein sim. GBank gjl114135[sp]P08205[JARGA_ECOLI - AMINO-ACID ACETYLTRANSFERASE (N-ACETYLGLUTAMATE SYNTHASE) (AGS)]		UNCLASSIFIED	265019
1098	79185424 (2195, 2196)	Novel Protein sim. GBank gjl114135[sp]P08205[JARGA_ECOLI - AMINO-ACID ACETYLTRANSFERASE (N-ACETYLGLUTAMATE SYNTHASE) (AGS)]			264687

1099	35523638 (2197, 2198)	Novel Protein sim. GBank gi 39151441sp O3307 TRMD_MYCLE - TRNA (GUANINE- NT)-METHYLTRANSFERASE (M1G- METHYLTRANSFERASE) (TRNA [GM37] METHYLTRANSFERASE)		UNCLASSIFIED	264603
1100	85736571 (2199, 2200)	Novel Protein sim. GBank gi 30232551sp O64420 ACOD_MESAU - ACYL-COA DESATURASE (STEARYL-COA DESATURASE) (FATTY ACID DESATURASE) (DELTA9)-DESATURASE)	desaturase		264259, 264636
1101	80491657 (2201, 2202)	Novel Protein sim. GBank gi 11747351sp P43012 TOP1_HAEN - DNA TOPOISOMERASE I (OMEGA-PROTEIN) (RELAXING ENZYME) (UNWISTING ENZYME) (SWIVELASE)	isomerase Contains protein domain (PF01396) - Topoisomerase DNA binding C4 zinc finger		264769
1102	79777614 (2203, 2204)	Novel Protein sim. GBank gi 1906596 (U81788) - kinesin-73		UNCLASSIFIED	264910, 264909
1103	81897259 (2205, 2206)	[Drosophila melanogaster]	struct		264757
1104	95003115 (2207, 2208)	Novel Protein sim. GBank gi 2935448 (AF048976) - synaptic ras GTPase-activating protein p135 SynGAP [Rattus norvegicus]		UNCLASSIFIED	29331822, 21806754, 264555, 264556, 264558, 22279002
1105	80255121 (2209, 2210)				264566
1106	79314110 (2211, 2212)			UNCLASSIFIED	264555, 264389
1107	80470019 (2213, 2214)				264908, 264769
1108	80440818 (2215, 2216)	Novel Protein sim. GBank gi 11734211sp P43416 SECY_STRSC - PREPROTEIN TRANSLOCASE SECY SUBUNIT	transport		264907, 264510, 264511, 264600, 264602, 264605, 264768, 264789
1109	80064615 (2217, 2218)	Novel Protein sim. GBank gi 2995310 emb CAA18338  - (AL022268) putative ATP-dependent helicase [Streptomyces coelicolor]	helicase		264602, 264605, 264639
1110	80503554 (2219, 2220)				264908, 264593, 265010, 264601, 264603, 264604, 264605, 264682, 264769, 264693, 264636
1111	80071744 (2221, 2222)	Novel Protein sim. GBank gi 2622039 (AE000868) - type I resistidion modification system, subunit S [Methanobacterium thermoautotrophicum]			18108370, 264557
1112	95010088 (2223, 2224)				264908
1113	82456352 (2225, 2226)	Novel Protein sim. GBank gi 3218378 emb CAA18628  - (AL023862) putative oxidoreductase [Streptomyces coelicolor]	UNCLASSIFIED		264600, 264602, 264604, 264605, 264762, 264769, 264585
1114	14698014 (2227, 2228)	Novel Protein sim. GBank gi 1083428 pir SS4876 - NAD(P)+ transhydrogenase (B-specific) (EC 1.6.1.1) precursor - mouse	dehydrogenase		264636
1115	11765583 (2229, 2230)				264686
1116	79841152 (2231, 2232)			UNCLASSIFIED	264908

1117	95305465 (2233, 2234)	Novel Protein sim. GBank gji3255965[emb]CAA94089] - (Z70200) U5 snRNP-specific 200KD protein [Homo sapiens]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	18108392, 264488, 263394, 264489, 18108398, 56182575, 22278995, 22278998, 35696288, 56994075, 22278997, 22278998, 22278999, 264259, 29331822, 56182181, 29331824, 66714117, 29331825, 29331826, 60432289, 29331827, 29331828, 35696052, 33656970, 29146488, 264508, 264905, 264509, 264806, 264907, 264908, 66712502, 264909, 52644045, 56182435, 264510, 264511, 265008, 265007, 264512, 265008, 265009, 264910, 60170831, 264591, 264592, 264593, 60433356, 264594, 60433438, 264595, 55812036, 264596, 264758, 264759, 21908754, 33108954, 33657084, 265011, 87168559, 264600, 264601, 264602, 264603, 265017, 264604, 265018, 264605, 265019, 264760, 55811150, 264681, 264762, 18108351, 264448, 264682, 264764, 264683, 264288, 264369, 264684, 264685, 264766, 264787, 264886, 264687, 264768, 264789, 264688, 18108359, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 29148784, 35695917, 265020, 265021, 265022, 60170815, 264690, 264691, 33657023, 264692, 264693, 65274820, 33657109, 33657182, 27486261, 27486262, 27486264, 33657349, 27486265, 35695783, 264691
1118	79563328 (2235, 2236)		UNCLASSIFIED	264691
1119	79642483 (2237, 2238)		UNCLASSIFIED	264907
1120	79480463 (2239, 2240)	Novel Protein sim. GBank gji5420387[emb]CAB46878.1] - (AJ243459) proteophosphoglycan [Leishmania major]	collagen	29331827, 265018, 265019, 264681, 265021, 60170615, 18108387
1121	79471716 (2241, 2242)	Novel Protein sim. GBank gji1644450 (U67864) - MEX-3 [Caenorhabditis elegans]	Contains protein domain (PF00013) - KH domain	264683, 264632, 18108388
1122	79456246 (2243, 2244)		UNCLASSIFIED	264639, 264563
1123	79637119 (2245, 2246)	Novel Protein sim. GBank gji98800[pir]S17768 - 3, dehydroquinase synthase (EC 4.6.1.3) - Mycobacterium tuberculosis	synthase	264693, 27488265
1124	79811596 (2247, 2248)		UNCLASSIFIED	264909
1125	79757861 (2249, 2250)		UNCLASSIFIED	264910
1126	79758914 (2251, 2252)	Novel Protein sim. GBank gji138154[sp]P03643[VG]_BPPHX - MAJOR SPIKE PROTEIN (G PROTEIN) (GPG)	eph	264905, 264809, 264910
1127	11800030 (2253, 2254)		UNCLASSIFIED	264682
1128	8364685 (2255, 2256)	Novel Protein sim. GBank gji500270[emb]CAB44358.1] - (AJ242630) DNA polymerase I [Methylobacterium sp. DM4]	Contains protein domain (PF00476) - DNA polymerase family A	264511

1129	80422480 (2257, 2258)	Novel Protein sim. GBank gij5689485[dbj EAA83026.1] - (AB028997) KIAA1074 protein [Homo sapiens]	Contains protein domain (PF00170) - bZIP transcription factor	UNCLASSIFIED	265011, 264766
1130	78420151 (2259, 2260)	Novel Protein sim. GBank		UNCLASSIFIED	264595
1131	80055391 (2261, 2262)	gi 4981328 gb AA035861.1 AE00174 - (AE00174) bioY protein [Thermotoga maritima]		UNCLASSIFIED	3589286, 22278998, 29331828, 264603, 264605, 264559
1132	82062248 (2263, 2264)	Novel Protein sim. GBank gij1841552 (U89336) - unknown [Homo sapiens]		UNCLASSIFIED	22278996, 264906, 265009, 264600, 264602, 264604, 264605, 264760, 32833986, 18108374
1133	17290437 (2265, 2266)			UNCLASSIFIED	265018
1134	80235376 (2267, 2268)			UNCLASSIFIED	264512, 264534
1135	80029393 (2269, 2270)	Novel Protein sim. GBank gij4539171[emb CA839700.1] - (AL049485) conserved hypothetical protein [Streptomyces coelicolor]		UNCLASSIFIED	264508, 264600, 264602, 264603, 18108376
1136	78842052 (2271, 2272)	Novel Protein sim. GBank gij4882454[gb AA036931.1 AE00182 - (AE00182) ATP-dependent protease LA, putative [Thermotoga maritima]		UNCLASSIFIED	264906, 264908
1137	90931557 (2273, 2274)	Novel Protein sim. GBank gij4972746[gb AA034768.1] - (AF132180) unknown [Drosophila melanogaster]	Contains protein domain (PF00515) - TPR Domain	collagen	22278998, 22278999, 35696052, 264907, 265009, 60433356, 264596, 265010, 264448, 264682, 264767, 264689, 265020, 264692, 55811578, 35685855, 284831, 264632, 22279002
1138	78841163 (2275, 2276)	Novel Protein sim. GBank gij731607[sp P30879 YHCA_YEAST - HYPOTHETICAL 63.8 KD PROTEIN IN GUT1-RIM1 INTERGENIC REGION PRECURSOR		struct	264908
1139	76633561 (2277, 2278)	Novel Protein sim. GBank gij3650031 (AC005396) - putative proline-rich cell wall protein [Arabidopsis thaliana]		UNCLASSIFIED	264693
1140	39480358 (2279, 2280)			UNCLASSIFIED	264593
1141	76638019 (2281, 2282)			UNCLASSIFIED	265019, 264693
1142	19835848 (2283, 2284)			UNCLASSIFIED	264631
1143	87782158 (2285, 2286)	Novel Protein sim. GBank gij3928000[emb CAA05880] - (AJ003125) procollagen I N-proteinase [Homo sapiens]	Contains protein domain (PF00090) - oxidase		56182575, 264908, 264600, 264632, 87168518
1144	80088988 (2287, 2288)		Thrombospondin type 1 domain	UNCLASSIFIED	264635, 264636, 264907, 264593, 264908, 264588, 264909
1145	14610262 (2289, 2290)			UNCLASSIFIED	264112
1146	82062092 (2291, 2292)			UNCLASSIFIED	264789, 284689, 35686286, 264760, 264805, 264486, 264559
1147	80071761 (2293, 2294)				264557
1148	80048433 (2295, 2296)	Novel Protein sim. GBank gij2499003[sp P76422 THID_ECOLI - PHOSPHOMETHYL-PYRIMIDINE KINASE (HMP-PHOSPHATE KINASE) (HMP-P KINASE)	kinase		264591
1149	11607438 (2297, 2298)	Novel Protein sim. GBank gij2896734[emb CAA17213.1] - (AL021897) hypothetical protein Rv1097c [Mycobacterium tuberculosis]			264591

1150	81325074 (2299, 2300)	Novel Protein sim. GBank gij2695065 (AF011337) - putative E1-E2 ATPase [Mus musculus]				264488, 35696286, 264907, 264908, 264909, 264910, 264593, 264596, 264758, 264764, 264766, 264768, 264693, 264628, 60431850, 264564, 264566, 264567
1151	80070874 (2301, 2302)	Novel Protein sim. GBank gij4324655jg AAD16978  (AF108191) DNA polymerase III alpha subunit [Streptomyces coelicolor]			polymerase	264595
1152	80235547 (2303, 2304)	Novel Protein sim. GBank gij3874275jemb CAB07311.1  - (292825) predicted using GeneFinder; Similarity to Yeast low affinity glucose transporter HXT4 (PS:32467); cDNA EST EMBL: C12555 comes from this gene; cDNA EST yk404c10.3 comes from this gene; cDNA EST yk404c10.5 comes from this...			glycoprotein	264488, 22278998, 264905, 264629, 264486
1153	80027783 (2305, 2306)	Novel Protein sim. GBank gij4240315jdb BAA74936.1  - (AB020720) KIAA0813 protein [Homo sapiens]				264910, 264555, 264557
1154	83002995 (2307, 2308)				UNCLASSIFIED	265008
1155	79411098 (2309, 2310)	Novel Protein sim. GBank gij58655jsp P37617 ATZN, ECOLI - ZINC-TRANSPORTING ATPASE [Zn(II)-TRANSLATING P-TYPE ATPASE]		Contains protein domain (PF00122) - E1-E2 ATPase	UNCLASSIFIED	264690, 264636
1156	57147843 (2311, 2312)	Novel Protein sim. GBank gij418480j P32139 YIHR, ECOLI - HYPOTHETICAL 34.0 KD PROTEIN IN GLNA-RBN INTERGENIC REGION			UNCLASSIFIED	264689, 264638, 264566
1157	95287711 (2313, 2314)	Novel Protein sim. GBank gij2488481j sp Q50724 Y09S_MYCTU - HYPOTHETICAL 87.3 KD PROTEIN CY78.27C				264906, 264762, 264687, 264769, 264689, 18108374, 35695855
1158	82454917 (2315, 2316)	Novel Protein sim. GBank gij1138408j BAA11490.1  - (D79995) similar to pig tubulin-tyrosine ligase. [Homo sapiens]				264687
1159	79186451 (2317, 2318)	Novel Protein sim. GBank gij2443342j BAA22380.1  - (D88764) alpha 2 type I collagen [Rana catesbeiana]				29331822, 29331824, 66714117, 29331826, 60433356, 265018, 265019, 83373044, 18108385, 22279000, 22279002, 264563
1160	91229893 (2319, 2320)				UNCLASSIFIED	264369
1161	7417143 (2321, 2322)	Novel Protein sim. GBank gij4503375j ref NP_001376.1 pDPYS - dihydropyrimidinase			UNCLASSIFIED	264693
1162	79635357 (2323, 2324)	Novel Protein sim. GBank gij50452554j BAA038607.1 AF14563 - (AF145632) BcDNA GH05032 [Drosophila melanogaster]			UNCLASSIFIED	29331827, 264906
1163	79553166 (2325, 2326)	Novel Protein sim. GBank gij50452554j BAA038607.1 AF14563 - (AF145632) BcDNA GH05032 [Drosophila melanogaster]			transport	18108398, 29331827, 29331828, 28146498, 29146499, 18108394, 21906768, 28148827, 21906769, 264693, 18108382, 18108385
1164	79650828 (2327, 2328)	Novel Protein sim. GBank gij50452554j BAA038607.1 AF14563 - (AF145632) BcDNA GH05032 [Drosophila melanogaster]			UNCLASSIFIED	264602, 264805, 264769, 18108370, 18108374, 264565
1165	80491888 (2329, 2330)					264488, 35696286, 22278999, 264259, 66714117, 60432289, 35898052, 264905, 56182435, 265006, 60433438, 264759, 21906754, 33109954, 265017, 265019, 264448, 264288, 264768, 264685, 35698423, 35695855, 264558, 18108385, 60432113
1166	88096456 (2331, 2332)	Novel Protein sim. GBank gij4589476j BAA76766.1  - (AB023139) KIAA0922 protein [Homo sapiens]				

1167	79863862 (2333, 2334)	Novel Protein sim. GBank gii2580433[dj]BAA23138] - [D76414] ppGpp hydrolase [Staphylococcus aureus]		kinase	264488
1168	8004678 (2335, 2336)			UNCLASSIFIED	264259, 28331827, 56182435, 60433438, 265019, 33657023, 35695855, 264566
1169	11805403 (2337, 2338)			UNCLASSIFIED	264581
1170	21632244 (2339, 2340)			UNCLASSIFIED	264502
1171	20434592 (2341, 2342)	Novel Protein sim. GBank gii272814 (AF028249) - precollagen D [Mytilus edulis]		UNCLASSIFIED	264556
1172	78610113 (2343, 2344)	Novel Protein sim. GBank gii4757848[re]NP_004317.1[pBCL8 - B-cell CLL/lymphoma 9]		UNCLASSIFIED	55810764, 35696052, 264634, 264488
1173	80235713 (2345, 2346)	Novel Protein sim. GBank gii2564053[dj]BAA22846] - (AB007832) Bm trachealless [Bombyx mori]			264508, 284806, 264907, 264909, 264591, 264632, 264638, 264639
1174	20293077 (2347, 2348)	Novel Protein sim. GBank gii2911027[emb]CAA17520] - (AL021958) mmsA [Mycobacterium tuberculosis]		dehydrogenase	264600
1175	20711847 (2349, 2350)	Novel Protein sim. GBank gii118333[sp]P23234[DCIP_ENTCL - INDOLE-3-PYRUVATE DECARBOXYLASE (INDOLEPYRUVATE DECARBOXYLASE)]		Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	264601
1176	80252845 (2351, 2352)	Novel Protein sim. GBank gii1144520 (U34956) - phosphoribosylformylglycinamide synthase [Mycobacterium tuberculosis]		Contains protein domain (PF00506) - AIR synthase related protein	264509, 264905, 264593, 264602, 264605
1177	80084647 (2353, 2354)	Novel Protein sim. GBank gii119791[sp]P28643[FBAG_CUPLA - 3-OXOACYL-(ACYL-CARRIER PROTEIN) REDUCTASE PRECURSOR (3-KETOACYL-ACYL CARRIER PROTEIN REDUCTASE)]		Contains protein domain (PF00106) - short chain dehydrogenase	264605
1178	94128641 (2355, 2356)	Novel Protein sim. GBank gii5031697[re]NP_005594.1[pFIC1 - familial intrahepatic cholestasis 1, (progressive, Byler disease and benign recurrent)]		Contains protein domain (PF00122) - E1-E2 ATPase	65274572, 18108398, 22278998, 22278999, 29331826, 264508, 264908, 264828, 33657402, 33109954, 264769, 21906765, 21906766, 21908788, 55811957, 33657023, 264629, 55811576, 35698423, 264638, 264556, 56182323, 60432113, 22279000, 22279002
1179	80055575 (2357, 2358)	Novel Protein sim. GBank gii2960090[emb]CAA17988.1] - (AL022121) dppA [Mycobacterium tuberculosis]		transport	264803
1180	11784446 (2359, 2360)	Novel Protein sim. GBank gii2558614[emb]CAA04787] - (AJ001493) dehydroquinase dehydratase [Streptomyces coelicolor]		Contains protein domain (PF00496) - Bacterial extracellular solute-binding proteins, family 5	264638
1181	17946362 (2361, 2362)	Novel Protein sim. GBank gii5420387[emb]CAB46678.1] - (AJ243459) proteophosphoglycan [Leishmania major]		Contains protein domain (PF01220) - Dehydroquinase class II	
1182	81494264 (2363, 2364)			UNCLASSIFIED	265017
1183	78574044 (2365, 2366)	Novel Protein sim. GBank gii4091877 (AF061331) - alpha galactosidase precursor [Saccharopolyspora erythraea]			265007, 265009, 264584, 264909, 264693
1184	52559933 (2367, 2368)	Novel Protein sim. GBank gii2126478[pr]IS51939 - chitinase (EC 3.2.1.4) precursor - beet		UNCLASSIFIED	264689, 35696423, 264638, 18108385
1185	79491185 (2369, 2370)			glycoprotein	264602
					263987

1186	20224012 (2371, 2372)				UNCLASSIFIED	264559	
1187	7924834 (2373, 2374)				UNCLASSIFIED	29331825, 265017, 18108351	
1188	79831387 (2375, 2376)	Novel Protein sim. GBank gij2998039 (AF054525) - hypothetical protein [Synecoccus PCC7002]			UNCLASSIFIED	264905, 264906	
1189	79609367 (2377, 2378)				UNCLASSIFIED	264692	
1190	78930589 (2379, 2380)				UNCLASSIFIED	285018	
1191	80310105 (2381, 2382)				UNCLASSIFIED	284600, 264605, 264764, 35695855, 264638,	
1192	13521841 (2383, 2384)				UNCLASSIFIED	264486	
1193	11103584 (2385, 2386)				UNCLASSIFIED	264636	
1194	78993947 (2387, 2388)	Novel Protein sim. GBank gij654065 (emb CAA58337  - (X83413) U88 [Human herpesvirus 6]			UNCLASSIFIED	263978	
1195	20445442 (2389, 2390)	Novel Protein sim. GBank gij1790277 (AE000459) - putative oxaloductase [Escherichia coli]			UNCLASSIFIED	265007, 265008	
1196	13000688 (2391, 2392)				UNCLASSIFIED	264605	
1197	11392317 (2393, 2394)	Novel Protein sim. GBank gij2497360 (sp Q50715 IMDH_MYCTU - INOSINE-5'- MONOPHOSPHATE DEHYDROGENASE (IMP DEHYDROGENASE) (IMPDH) (IMPD)			UNCLASSIFIED	264689	
1198	85280101 (2395, 2396)				UNCLASSIFIED	264594	
1199	81882011 (2397, 2398)	Novel Protein sim. GBank gij1708525 (sp P54673 P3K1_DICD1 - PHOSPHATIDYLINOSITOL 3-KINASE 1 (PI3-KINASE) (PTDINS-3-KINASE) (PI3K)			UNCLASSIFIED	264603	
1200	9846860 (2399, 2400)				UNCLASSIFIED	284259, 264757, 33109954, 21908768	
1201	80503751 (2401, 2402)	Novel Protein sim. GBank gij2498877 (sp P70845 BLMH_RAT - BLEOMYCIN HYDROLASE (BLM HYDROLASE) (BMH)			UNCLASSIFIED	264910	
1202	80082633 (2403, 2404)	Novel Protein sim. GBank gij606342 (U18997) - ORF_c622; reading frame open far upstream of start; possible frameshift, linking to previous ORF [Escherichia coli]			UNCLASSIFIED	264766, 264769	
1203	82125373 (2405, 2406)				UNCLASSIFIED	264600, 264558	
1204	80503916 (2407, 2408)	Novel Protein sim. GBank gij2500728 (sp Q59812 SECY_STRGB - PREPROTEIN TRANSLUCASE SECY SUBUNIT			UNCLASSIFIED	264766, 264769, 35695917, 264910, 264760, 264906, 264907, 264829, 264908, 264909, 264766	
1205	80053981 (2409, 2410)				UNCLASSIFIED	264905, 264769, 264636	
1206	80241885 (2411, 2412)				UNCLASSIFIED	264566	
1207	79841182 (2413, 2414)				UNCLASSIFIED	264556, 264557, 264558	
1208	87755217 (2415, 2416)	Novel Protein sim. GBank gij2645560 (AF027954) - Bcl-2- related ovarian killer protein [Rattus norvegicus]			UNCLASSIFIED	29331824, 264909, 265021, 18108370	
					UNCLASSIFIED	29331824, 29331825, 29331827, 265007, 264764, 264683, 264769, 264688, 264689	



1209	79185742 (2417, 2418)	Novel Protein sim. GBank gij1175033[sp]P44398[X]YLA_HAEIN - XYLOSE ISOMERASE	Contains protein domain (PF00259) - Xylose isomerase	isomerase	264687, 264688
1210	56426884 (2419, 2420)			UNCLASSIFIED	264907, 264693
1211	94665655 (2421, 2422)	Novel Protein sim. GBank gij421095[pr]j530688 - hypothetical protein g248 - Escherichia coli		transferase	264591, 264592, 264595
1212	79167929 (2423, 2424)	Novel Protein sim. GBank gij3880625[emb]CAB07858] - (Z93785) predicted using GeneFinder; similar to RNA recognition motif. (aka RRM, RBD, or RNP domain); cDNA EST EMBL.T01882 comes from this gene; cDNA EST EMBL.M75823 comes from this gene; cDNA EST EMBL.D27559 comes from this ge...	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf		264689, 263967
1213	79859633 (2425, 2426)	Novel Protein sim. GBank gij228282[pr]j1505375A - vir gene [Bordetella pertussis]		kinase	264909
1214	10144306 (2427, 2428)	Novel Protein sim. GBank gij5726285[g]AAD48396.1[AF12616] - (AF126162) HERV-H LTR associating protein 2 [Homo sapiens]		UNCLASSIFIED	264908
1215	80050106 (2429, 2430)	Novel Protein sim. GBank gij232679[emb]CAB10953] - (Z98269) recN [Mycobacterium tuberculosis]		UNCLASSIFIED	265009, 264601, 264602, 264603, 33657109
1216	20438324 (2431, 2432)	Novel Protein sim. GBank gij417329[sp]P33038[MURA_ENTCL - UDP-N- ACETYLGLUCOSAMINE 1- CARBOXYVINYLTRANSFERASE (ENOYLPIRUVATE TRANSFERASE) (UDP-N-ACETYLGLUCOSAMINE ENOYLPIRUVYL TRANSFERASE) (EPT)		transferase	264604
1217	95011344 (2433, 2434)			UNCLASSIFIED	264905, 264907, 264908, 264591, 264766, 264681, 264693, 264628, 264630, 264636, 264584
1218	11063860 (2435, 2436)	Novel Protein sim. GBank gij1805460[db]BAA09022] - (D50453) homologue of succinate semialdehyde dehydrogenase GabD of E. coli [Bacillus subtilis]		dehydrogenase	264601
1219	91216252 (2437, 2438)				56181686, 29331822, 60432289, 264601, 264692, 264629
1220	91241524 (2439, 2440)	Novel Protein sim. GBank gij4240315[db]BAA74936.1] - (AB020720) KIAA0913 protein [Homo sapiens]		oncogene	52644507, 264905, 264909, 265008, 265019, 265020, 52644150, 33657023, 264693, 33657182, 35695763, 264634, 22279000, 22279002, 264482
1221	83045055 (2441, 2442)	Novel Protein sim. GBank gij2143886[pr]j52523 - nucleoporin p82 homologue - rat (fragment)		UNCLASSIFIED	264768, 265020, 264906
1222	20711865 (2443, 2444)	Novel Protein sim. GBank gij730807[sp]P39663[SPHR_SYN7 - ALKALINE PHOSPHATASE SYNTHESIS TRANSCRIPTIONAL REGULATORY PROTEIN SPHR	Contains protein domain (PF00486) - Transcriptional regulatory protein, C terminal	phosphatase	264601
1223	11815847 (2445, 2446)				264593
1224	80432845 (2447, 2448)	Novel Protein sim. GBank gij1172627[sp]P46546[PROB_CORGL - GLUTAMATE 5- KINASE (GAMMA-GLUTAMYL KINASE) (GK)	Contains protein domain (PF01472) - PUA domain	kinase	264593, 264600, 264601, 264603, 264605, 264768, 18108376, 264635, 18108387

1225	80434427 (2449, 2450)	Novel Protein sim. GBank gij2105050[emb]CAB08836] - (Z95436) hypothetical protein Rv3644c [Mycobacterium tuberculosis]			polymerase	264768 264905, 264512, 264689
1226	80237518 (2451, 2452)	Novel Protein sim. GBank gij1706769[sp]P98133[FBN1_BOVIN - FIBRILLIN 1 PRECURSOR (MP340)]			UNCLASSIFIED	264808, 264637, 264639
1227	79422138 (2453, 2454)	Novel Protein sim. GBank gij1706769[sp]P98133[FBN1_BOVIN - FIBRILLIN 1 PRECURSOR (MP340)]				264605, 264634
1228	78209027 (2455, 2456)	Novel Protein sim. GBank gij165390[db]JBA118811] - (D90817) acriflavine resistance protein [Synectocystis sp.]			UNCLASSIFIED	87168474, 265011, 87168559, 264681, 264689, 264693, 65274620, 18108374
1229	94329135 (2457, 2458)	Novel Protein sim. GBank gij165390[db]JBA118811] - (D90817) acriflavine resistance protein [Synectocystis sp.]			UNCLASSIFIED	264605, 264634
1230	80049357 (2459, 2460)	Novel Protein sim. GBank gij116230[sp]P28598[CH60_BACSU - 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN)]			UNCLASSIFIED	87168474, 265011, 87168559, 264681, 264689, 264693, 65274620, 18108374
1231	79843141 (2461, 2462)	Novel Protein sim. GBank gij116230[sp]P28598[CH60_BACSU - 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN)]			UNCLASSIFIED	264605, 264634
1232	79853104 (2463, 2464)	Novel Protein sim. GBank gij116230[sp]P28598[CH60_BACSU - 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN)]			UNCLASSIFIED	264605, 264634
1233	80255179 (2465, 2466)	Novel Protein sim. GBank gij1215733 (U48718) - OphC [Agrobacterium tumefaciens]			UNCLASSIFIED	264908 264909
1234	78242158 (2467, 2468)	Novel Protein sim. GBank gij1215733 (U48718) - OphC [Agrobacterium tumefaciens]			UNCLASSIFIED	265017, 264564
1235	79914423 (2469, 2470)	Novel Protein sim. GBank gij1215733 (U48718) - OphC [Agrobacterium tumefaciens]			UNCLASSIFIED	265008, 265010, 18108381
1236	81927147 (2471, 2472)	Novel Protein sim. GBank gij1215733 (U48718) - OphC [Agrobacterium tumefaciens]			UNCLASSIFIED	264634, 264762
1237	83371782 (2473, 2474)	Novel Protein sim. GBank gij1215733 (U48718) - OphC [Agrobacterium tumefaciens]			UNCLASSIFIED	265018, 55811150, 264585, 264757 264758, 264601, 264766, 264697, 18108372, 264555, 284559
1238	87411577 (2475, 2476)	Novel Protein sim. GBank gij3875133[emb]CAA94750] - (Z70750) similar to actin binding domain; cDNA EST EMBL:Y00093 comes from this gene; cDNA EST EMBL:D34443 comes from this gene; cDNA EST EMBL:D37508 comes from this gene; cDNA EST EMBL:D84247 comes from this gene; cDNA EST EMBL: protein-coupled receptor LGR4 [Rattus norvegicus]			glycoprotein	264259, 29331822, 29331824, 35696052, 264508, 264906, 52644045, 52648317, 264288, 264769, 264693, 264632, 264634, 264558, 87168518, 264563
1239	82197449 (2477, 2478)	Novel Protein sim. GBank gij4007990[gb]AAC95339] - (AF084383) DOK protein [Mus musculus]			oncogene	264508, 284511, 264759, 264760, 264764, 264557
1240	80497259 (2479, 2480)	Novel Protein sim. GBank gij1176192[sp]P45420[YHCO_ECOLI - HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN IN GLTF-NANT INTERGENIC REGION PRECURSOR]				264769
1241	80020711 (2481, 2482)	Novel Protein sim. GBank gij121383[sp]P19804[GLNA_VIBAL - GLUTAMINE SYNTHETASE (GLUTAMATE-AMMONIA LIGASE)]			UNCLASSIFIED	264601, 264604, 264638
1242	79775890 (2483, 2484)	Novel Protein sim. GBank gij121383[sp]P19804[GLNA_VIBAL - GLUTAMINE SYNTHETASE (GLUTAMATE-AMMONIA LIGASE)]				264906, 264807, 264808, 264634

1243	79779458 (2485, 2486)	Novel Protein sim. GBank gij3356671emb[CAA19971] - (AL031124) branched-chain amino acid aminotransferase [Streptomyces coelicolor]		UNCLASSIFIED	18108374, 35895917, 35895855, 265009, 264508, 264908
1244	10284821 (2487, 2488)	Novel Protein sim. GBank gij2970646 (AF051945) - Xin [Mus musculus]		UNCLASSIFIED	264691
1245	80437103 (2489, 2490)	Novel Protein sim. GBank gij4586338[dbj]BAA76357.1 - (AB016787) cytochrome o ubiquinol oxidase B [Pseudomonas putida]	Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	oxidase	264768
1246	80059321 (2491, 2492)	Novel Protein sim. GBank gij3581849[emb]CAA20805] - (AL031541) putative phenylalanine-tRNA synthetase beta chain [Streptomyces coelicolor]		UNCLASSIFIED	264604, 264636, 264557, 264564
1247	80064831 (2493, 2494)	Novel Protein sim. GBank gij2621884 (AE000842) - adhesion protein [Methanobacterium thermoautotrophicum]			264758, 264605, 264639
1248	38070353 (2495, 2496)	Novel Protein sim. GBank gij1352403[sp]P09467IF16P_HUMAN - FRUCTOSE-1,6-BISPHOSPHATASE (D-FRUCTOSE-1,6-BISPHOSPHATE 1-PHOSPHOHYDROLASE) (FBPASE)	Contains protein domain (PF00316) - Fructose-1,6-bisphosphatase	UNCLASSIFIED	18108392, 264259, 29331826, 264106, 264508, 264907, 264828, 265009, 60433356, 264757, 264758, 21008754, 265010, 265011, 265018, 265019, 264780, 18108351, 18108354, 265021, 18108376, 18108377, 264630, 18108385
1249	30055657 (2497, 2498)	Novel Protein sim. GBank gij2791407[emb]CAA16001] - (AL021184) hypothetical protein RV1473 [Mycobacterium tuberculosis]		transport	264908, 265010, 264600, 264603, 264691, 18108376
1250	12684385 (2499, 2500)	Novel Protein sim. GBank gij112785[sp]P051003MG1_ECOLI - DNA-3-METHYLADENINE GLYCOSIDASE I (3-METHYLADENINE-DNA GLYCOSYLASE I, CONSTITUTIVE) (TAG I)		UNCLASSIFIED	264689
1251	79850448 (2501, 2502)				264809
1252	79458087 (2503, 2504)			UNCLASSIFIED	264583, 263976
1253	80050121 (2505, 2506)	Novel Protein sim. GBank gij5670176[gb]AAD46616.1[AF16131] - (AF16131) NRAMP manganese transport protein MntA [Salmonella typhimurium]		glycoprotein	264600, 264603, 18108376
1254	87718767 (2507, 2508)	Novel Protein sim. GBank gij103160[pir]J522125 - finger protein unkempt - fruit fly (Drosophila melanogaster)		UNCLASSIFIED	35896286, 264910, 264784, 264688, 21906767, 55811957, 264692, 264556, 264639
1255	79169728 (2509, 2510)				264636
1256	87889508 (2511, 2512)	Novel Protein sim. GBank gij2995353[emb]CAA04608.1] - (AJ001206) pep2 [Streptomyces coelicolor]		UNCLASSIFIED	60432289, 264600, 264605, 264764, 264687, 264789, 264689, 27486285, 18108374, 18108376
1257	80201435 (2513, 2514)	Novel Protein sim. GBank gij3193306 (AF068300) - contains similarity to Arabidopsis membrane-associated salt inducible-like protein (GB:AL021637) [Arabidopsis thaliana]		UNCLASSIFIED	264094, 265019
1258	20708150 (2515, 2516)			UNCLASSIFIED	264602, 263978
1259	80186012 (2517, 2518)			UNCLASSIFIED	264906, 264448, 264908
1260	80084808 (2519, 2520)			UNCLASSIFIED	264634, 264639

1261	87412802 (2521, 2522)	Novel Protein sim. GBank gij568951jdbjBAA83039.1] - (AB028010) KIAA1087 protein [Homo sapiens]	Contains protein domain (PF01698) - Sodium/calcium exchanger protein	cadherin	28331824, 284908, 284909, 284768, 284769, 284689, 284693, 284639, 18108384, 284563, 284634
1262	13304589 (2523, 2524)	Novel Protein sim. GBank gij95100jpirjS21334 - hypothetical protein 4 - Agrobacterium tumefaciens		UNCLASSIFIED	
1263	20710897 (2525, 2526)	Novel Protein sim. GBank gij3550958 (AF004840) - CDO [Rattus norvegicus]		struct	284602, 284634
1264	80083396 (2527, 2528)	Novel Protein sim. GBank gij1085002jpirjS55055 - mitochondrial carrier protein DIF-1 homolog - Caenorhabditis elegans	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	284583
1265	80253578 (2529, 2530)	Novel Protein sim. GBank gij4886445jmbjCAB43370.1] - (AL050286) hypothetical protein [Homo sapiens]		UNCLASSIFIED	284768, 284638, 284638, 284567
1266	79914604 (2531, 2532)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	transport	284259, 21808754, 264369
1267	80558918 (2533, 2534)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	18108398, 22278995, 56994075, 60424269, 28331827, 284109, 264512, 265007, 265008, 265009, 264595, 33109934, 33657084, 87168559, 284600, 265018, 265018, 264369, 284688, 21808767, 265020, 52844150, 284691, 33657023, 33657349, 18108374, 264556, 18108385, 60432113, 22279002, 264486
1268	88178473 (2535, 2536)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1269	79821948 (2537, 2538)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1270	80031420 (2539, 2540)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1271	79840498 (2541, 2542)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1272	79462878 (2543, 2544)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1273	80220315 (2545, 2546)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1274	95010802 (2547, 2548)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1275	20730763 (2549, 2550)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1276	21148644 (2551, 2552)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1277	20438195 (2553, 2554)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1278	11088365 (2555, 2556)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1279	21658756 (2557, 2558)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693
1280	79310959 (2559, 2560)	Novel Protein sim. GBank gij1655665jmbjCAB03731] - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]	Contains protein domain (PF01574) - dehydrogenase	UNCLASSIFIED	264508, 264905, 264906, 284687, 264693

1281	94323988 (2561, 2562)	Novel Protein sim. GBank gll1136501 (U39546) - surface protein MCA-32 [Rattus norvegicus]	Contains protein domain (PF00047) - Immunoglobulin domain	UNCLASSIFIED	29331825, 29331828, 264766, 83373044
1282	87537895 (2563, 2564)	Novel Protein sim. GBank gll3328180 (AF074266) - proto-oncogene AF4 [Mus musculus]		UNCLASSIFIED	265008
1283	20466305 (2565, 2566)	Novel Protein sim. GBank gll326172 (emb) (CAB07057) - (Z92770) hypothetical protein Rv0153c [Mycobacterium tuberculosis]		UNCLASSIFIED	284605
1284	20636325 (2567, 2568)	Novel Protein sim. GBank gll3929022 (AF057686) - LspB [Haemophilus ducreyi]			284604
1285	80427330 (2569, 2570)	Novel Protein sim. GBank gll417154 (sp) P33126 (H) S82_ORYSA - HEAT SHOCK PROTEIN 82 (Z55207) gorA [Mycobacterium tuberculosis]	Contains protein domain (PF00183) - eph		264766, 264689, 263967
1286	20465254 (2571, 2572)	Novel Protein sim. GBank gll2078004 (emb) (CAB08451) - (Z55207) gorA [Mycobacterium tuberculosis]	reductase		264605, 264639
1287	80417530 (2573, 2574)			UNCLASSIFIED	265011, 264602, 264766, 264687, 264769, 264689, 18108370, 264636, 18108385, 264583
1288	95338101 (2575, 2576)	Novel Protein sim. GBank gll5353510 (gb) IAD42161.1 (AF08891) - (AF088916) emilin precursor [Homo sapiens]	Contains protein domain (PF00386) - collagen C1q domain		35695032, 264107, 264508, 264509, 264905, 264908, 264907, 264908, 264909, 264510, 264511, 265007, 264512, 284910, 265009, 33657402, 264595, 264758, 265011, 285019, 264760, 18108351, 264681, 264764, 264288, 264685, 264768, 264887, 264768, 264769, 265020, 285021, 264534, 264692, 18108370, 264628, 18108374, 35886423, 264555, 264556, 264557, 264558, 18108385, 264564, 264586, 264587, 264486, 18108391
1289	11813647 (2577, 2578)	Novel Protein sim. GBank gll1169985 (sp) P46023 (GPCR_LYMST - G-PROTEIN COUPLED RECEPTOR GRL101 PRECURSOR (Z95120) rhIE [Mycobacterium tuberculosis]	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	UNCLASSIFIED	264637
1290	19526027 (2578, 2580)	Novel Protein sim. GBank gll2072674 (emb) (CAB08305) - (Z95120) rhIE [Mycobacterium tuberculosis]	Contains protein domain (PF00271) - helicase		264583
1291	80470266 (2581, 2582)	Novel Protein sim. GBank gll1835755 (U86338) - zinc finger protein Png-1 [Mus musculus]	Helicases conserved C-terminal domain	265007, 265008, 264769	
1292	94723316 (2583, 2584)	Novel Protein sim. GBank gll1835755 (U86338) - zinc finger protein Png-1 [Mus musculus]	Contains protein domain (PF01530) - Zinc finger, C2HC type		264092, 264259, 29331822, 29331824, 264508, 264908, 264909, 264512, 265008, 265009, 264591, 285019, 264369, 264288, 264686, 264768, 264693, 18108374, 264632, 56182323, 264639, 83373044, 22279002, 264482, 264563
1293	80067538 (2585, 2586)	Novel Protein sim. GBank gll2129173 (pr) (F8453 - oxalacetate decarboxylase (EC 4.1.1.3) alpha subunit - Methanococcus jannaschii)		UNCLASSIFIED	265006, 55812038, 264369, 264556
1294	82125908 (2587, 2588)	Novel Protein sim. GBank gll5441779 (emb) (CAB46803.1) - (AL096811) putative alcohol dehydrogenase (zinc-binding) [Streptomyces coelicolor A3(2)]	biotindep		264602, 264605, 264760, 18108351, 264689, 33657023, 264559
1295	11685651 (2589, 2590)	Novel Protein sim. GBank gll5441779 (emb) (CAB46803.1) - (AL096811) putative alcohol dehydrogenase (zinc-binding) [Streptomyces coelicolor A3(2)]	dehydrogenase		264689

1296	11687904 (2591, 2592)	Novel Protein sim. GBank g 492191 gb AAD35686.1 AE00180 - (AE00180) DNA-directed DNA polymerase I [Thermotoga maritima]	Contains protein domain (PF01367) - polymerase 5'-3' exonuclease	UNCLASSIFIED	264591, 264639
1297	78639300 (2593, 2594)	Novel Protein sim. GBank g 492191 gb AAD35686.1 AE00180 - (AE00180) DNA-directed DNA polymerase I [Thermotoga maritima]			264693
1298	94239508 (2595, 2596)	Novel Protein sim. GBank g 1943770 (U97191) - F53F10.1 gene product [Caenorhabditis elegans]		struct	18108348, 265017
1299	80255378 (2597, 2598)	Novel Protein sim. GBank g 3445181 (AC005498) - R31655_2 [Homo sapiens]			264488, 264906, 264909, 22279002, 264566
1300	80064867 (2599, 2600)	Novel Protein sim. GBank g 4082973 gb BAA36204.1  - (AB017138) alpha subunit of malonate decarboxylase [Pseudomonas putida]	Contains protein domain (PF01352) - KRAB box	transcript lador	264605
1301	17839614 (2601, 2602)	Novel Protein sim. GBank g 3242273 emb CAB070171 - (Z92668) hypothetical protein Rv0238c [Mycobacterium tuberculosis]		UNCLASSIFIED	264806
1302	95416198 (2603, 2604)				85658542, 265020
1303	9684121 (2605, 2606)				264908
1304	79377196 (2607, 2608)			UNCLASSIFIED	264508
1305	19905699 (2609, 2610)				264566
1306	13069230 (2611, 2612)			UNCLASSIFIED	264636
1307	82201028 (2613, 2614)	Novel Protein sim. GBank g 1502421 (U59433) - 3-ketolacyl acyl carrier protein reductase [Bacillus subtilis]	Contains protein domain (PF00516) - reductase	UNCLASSIFIED	264907, 264592, 264784
1308	21428814 (2615, 2616)	Novel Protein sim. GBank g 85819 pir S16298 - ferric enterobactin transport protein lepC - Escherichia coli	Envelope glycoprotein GP120	264535	
1309	79263011 (2617, 2618)	Novel Protein sim. GBank g 5459220 emb CAB48893.1  - (AL096837) putative iron-sulfur protein [Streptomyces coelicolor A3(2)]	transport		264906, 18108354
1310	20468319 (2619, 2620)	Novel Protein sim. GBank g 5459220 emb CAB48893.1  - (AL096837) putative iron-sulfur protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	264805
1311	87613142 (2621, 2622)				35696286, 29331827, 264908, 265008, 264764, 264766, 264686, 21909787, 21906789, 35695917, 264691, 264693, 22278995, 22278996, 22278999, 264905, 264808, 265011, 265017, 265019, 264687, 21906788, 265020, 265021, 33657023, 22279002, 264584
1312	88081720 (2623, 2624)	Novel Protein sim. GBank g 4455118 gb AAD210841 - (AF125158) zinc finger DNA binding protein 99 [Homo sapiens]	Contains protein domain (PF00096) - dna_ma_bind Zinc finger, C2H2 type		22278996, 22278999, 264259, 20281089, 29146496, 264508, 264908, 68712502, 60433356, 60433438, 265011, 265017, 264683, 264298, 21906765, 21906767, 29148827, 21906768, 35695917, 265021, 33657023, 33657109, 18108370, 18108377, 35695855, 60432113, 22279000, 264563, 18108390
1313	91225458 (2625, 2626)	Novel Protein sim. GBank g 4928733 gb AAD34127.1 AF15189 - (AF15189) CGI-132 protein [Homo sapiens]	Contains protein domain (PF00886) - ribosomalprot Ribosomal protein S16		
1314	56926053 (2627, 2628)				264693
1315	84357192 (2629, 2630)	Novel Protein sim. GBank g 2589223 (AF026565) - ring finger protein [Mus musculus]	Contains protein domain (PF00097) - ring Zinc finger, C3HC4 type (RING finger)	interleukinrecept	264691

1316	95361609 (2631, 2632)	Novel Protein sim. GBank gij5689407[dbj BAA02987.1] - (AB028958) KIAA1035 protein [Homo sapiens]		kinase	56182575, 56181686, 20281171, 29331822, 29331824, 60424269, 29331825, 35696052, 52644045, 264591, 60432229, 265018, 265019, 55811150, 56181562, 21906765, 21906767, 21906768, 35695917, 60170815, 33657023, 65274620, 33657109, 35695763, 35695855, 18108387, 87168518, 60432113, 22279002, 264594
1317	98055167 (2633, 2634)	Novel Protein sim. GBank gij4836757[gb AAD30541.1 AF13491 - (AF134918) semaphorin subclass 4 member G [Mus musculus]		UNCLASSIFIED	264093, 264906, 264909, 264369, 264684
1318	95322893 (2635, 2636)	Novel Protein sim. GBank gij4680204[gb AAD27587.1 AF11417 - (AF114171) hypothetical protein [Sorghum bicolor]		UNCLASSIFIED	18108392, 18108348, 265011, 265017, 18108359, 18108382, 58182323, 18108385, 22279000
1319	94238546 (2637, 2638)				264908, 264909, 265006, 265008, 264592, 265019, 264766, 56181562, 18108388, 264628, 264629, 18108377, 264636
1320	86603567 (2639, 2640)	Novel Protein sim. GBank gij4240183[dbj BAA74870.1] - (AB020854) KIAA0847 protein [Homo sapiens]		UNCLASSIFIED	35696288, 59812038, 265018, 21906768, 265020, 263978, 22279002
1321	86676351 (2641, 2642)	Novel Protein sim. GBank gij4886503[emb CAB43377.1] - (AL050276) hypothetical protein [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	transcriptfactor	60432049, 28331828, 264907, 264908, 264909, 264910, 55812038, 264601, 264762, 264764, 264766, 264768, 264769, 264628, 18108374, 264634, 264635, 18108385
1322	87755272 (2643, 2644)	Novel Protein sim. GBank gij5262591[emb CAB45736.1] - (AL080143) hypothetical protein [Homo sapiens]	Contains protein domain (PF00095) - Zinc finger, C2H2 type	dna_mg_bind	29331828, 264908, 265020, 33657023, 264693, 264404
1323	94845931 (2645, 2646)	Novel Protein sim. GBank gij545951[dbj BAA82407.1] - (AB029821) phosphatidylethanolamine N-methyltransferase [Homo sapiens]		synthase	65274572, 56994075, 264259, 29331822, 29331827, 264104, 56182435, 87188474, 18108351, 264288, 21906766, 21906767, 35695917, 265020, 264693, 64274791, 58182323, 18108387
1324	87737614 (2647, 2648)	Novel Protein sim. GBank gij5031717[ref NP_005704.1 pGPBP - goodpasture antigen-binding protein	Contains protein domain (PF01852) - START domain		22278986, 22278988, 29331828, 264905, 264907, 29331830, 264908, 264510, 265008, 264595, 264759, 21906754, 265018, 264288, 264768, 264769, 21906766, 265022, 18108378, 264831, 264832, 264834, 264836
1325	94847471 (2649, 2650)	Novel Protein sim. GBank gij3294501 (U64857) - similar to the DPT/Kunitz family of inhibitors; most similar to tissue factor pathway inhibitor precursor [Caenorhabditis elegans]	Contains protein domain (PF00090) - Thrombospondin type 1 domain	protease	264638, 264563, 264564, 264565, 264566, 35696288, 264905, 264906, 264907, 264908, 264909, 264910, 264593, 33657402, 264758, 85658542, 284760, 284768, 264769, 264691, 35696423
1326	87316289 (2651, 2652)	Novel Protein sim. GBank gij1397275 (U61947) - C06G3.8 gene product [Caenorhabditis elegans]		UNCLASSIFIED	264259, 66712502, 264682, 264683, 264635

1327	95322897 (2653, 2654)	Novel Protein sim. GBank gij728832(splP39189)ALU2_HUMAN - III ALU SUBFAMILY SB WARNING ENTRY IIII	Contains protein domain (PF00279) - Plant lipid transfer protein family	UNCLASSIFIED	18108398, 22278996, 22278997, 22278999, 264091, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264105, 264905, 56182435, 264112, 265008, 265009, 21908754, 265010, 265011, 265017, 265019, 264661, 264448, 264764, 264664, 264288, 264665, 264768, 264688, 21906767, 21906769, 26148629, 265020, 264660, 264691, 264692, 264693, 263967, 33657109, 33657182, 27486282, 33657348, 18108370, 18108374, 55810764, 35695855, 264634, 56182323, 63373044, 87168518, 60432113, 22279000, 22279002, 264583, 264587
1328	87753493 (2655, 2656)			UNCLASSIFIED	264488, 22276997, 29331828, 264595, 18108351, 264766, 22279002, 284482, 264567
1329	87755276 (2657, 2658)	Novel Protein sim. GBank gij4678224(gbIAAD26969.1)AC00713 - (AC007135) unknown protein (Arabidopsis thaliana)		UNCLASSIFIED	22278996, 29331827, 264684, 264692, 33657109
1330	87727737 (2658, 2660)	Novel Protein sim. GBank gij437310 (L23504) - nodulin [Medicago truncatula]		UNCLASSIFIED	264259, 29331825, 264512, 265019, 265021, 264555, 264558, 56526486
1331	87376764 (2661, 2662)	Novel Protein sim. GBank gij4589586(dbjBAA76815.1) - (AB023188) KIAA0971 protein [Homo sapiens]		UNCLASSIFIED	264259, 29331826, 29331827, 35686052, 29331828, 60170831, 264448, 264686, 21906765, 55811957, 265020, 33657023, 33657109, 263973, 55811576, 35696423, 35695655, 56182323
1332	94845937 (2663, 2664)	Novel Protein sim. GBank gij5459516(dbjBAA82407.1) - (AB029821) phosphatidylethanolamine N-methyltransferase [Homo sapiens]		synthase	65274572, 22278996, 56994075, 22278999, 60432049, 264259, 29331822, 29331826, 60432288, 29331827, 35696052, 52844045, 56182435, 264510, 21906754, 87168559, 265018, 265019, 264448, 264288, 264369, 264686, 21906765, 21906766, 21906767, 21906768, 265020, 265021, 52844150, 33657023, 56182323, 18108387, 60432113, 22279002
1333	88098476 (2665, 2666)	Novel Protein sim. GBank gij5689527(dbjBAA83047.1) - (AB029018) KIAA1095 protein [Homo sapiens]		UNCLASSIFIED	60432289, 66712502, 264591, 60433356, 60433438, 55812038, 265010, 264639, 56526488
1334	87592388 (2667, 2668)	Novel Protein sim. GBank gij2662536 (AF036665) - Similar to protein-tyrosine phosphatase [Caenorhabditis elegans]		phosphatase	264605
1335	87644798 (2668, 2670)	Novel Protein sim. GBank gij4240285(dbjBAA74921.1) - (AB020705) KIAA0888 protein [Homo sapiens]	Contains protein domain (PF00843) - B-box zinc finger.	UNCLASSIFIED	22278996, 22278999, 29331827, 264509, 264511, 265007, 265008, 265009, 60433438, 21906754, 87168559, 265017, 264288, 21906765, 21906767, 21906768, 21906769, 265020, 33657108, 27486284, 18108374, 264558, 264638, 264557, 60170394, 264559, 18108395, 264563



1336	87787890 (2671, 2672)	Novel Protein sim. GBank gi 465445 sp P33485 VNUA_PRIVKA - PROBABLE NUCLEAR ANTIGEN			UNCLASSIFIED	264509, 264905, 264512, 264764, 264693, 264635, 264637
1337	94312042 (2673, 2674)	Novel Protein sim. GBank gi 5669471 dbj BAA83019.1  - (AB028990) KIAA1067 protein [Homo sapiens]				56182575, 56994075, 22278998, 22278999, 264092, 264259, 60432289, 29331828, 264908, 264908, 264908, 264112, 265008, 285009, 60433358, 55812038, 33657084, 265011, 265017, 265018, 265019, 264682, 264448, 264683, 264369, 264688, 264689, 21808766, 21808769, 265020, 264681, 27486261, 20281069, 18108379, 55811578, 35695855, 56182323, 60432113, 22279002, 264567
1338	80366114 (2675, 2676)				UNCLASSIFIED	29331822, 265010, 264288, 264689, 18108370, 35695855
1339	80249231 (2677, 2678)	Novel Protein sim. GBank gi 176422 (U43194) - rhophilin [Mus musculus]			UNCLASSIFIED	35696032, 264909, 264688, 264556, 264558
1340	88316311 (2679, 2680)					264905, 264907, 87188559, 264784
1341	88101485 (2681, 2682)					264681, 264685, 264688, 264692
1342	80089017 (2683, 2684)	Novel Protein sim. GBank gi 5019564 emb CAB44507.1  - (AL035542) dJ994E9.5 (hs6M1-17 (novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like) protein)) [Homo sapiens]	Contains protein domain (PF00001) - tm7 7 transmembrane receptor (rhodopsin family)			264629
1343	80082662 (2685, 2686)	Novel Protein sim. GBank gi 4557543 ref NP_001384.1 pECM2 - extracellular matrix protein 2	Contains protein domain (PF00560) - struct Leucine Rich Repeat			264910, 264688, 264634
1344	20562558 (2687, 2688)	Novel Protein sim. GBank gi 214410 ipr I55210 - Itricarboxylate carrier - rat (fragment)			glycoprotein	263978
1345	91225546 (2689, 2690)	Novel Protein sim. GBank gi 3881052 emb CAA19523  - (AL023843) predicted using GeneFinder: similar to serine/threonine kinase; cDNA EST yk248a12.3 comes from this gene; cDNA EST yk359c10.5 comes from this gene; cDNA EST EMBL:M89047 comes from this gene; cDNA EST yk248a12.5 comes...	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain			22278998, 264907, 264681, 264685, 264689, 265020, 264693, 22279000, 22279002, 264568
1347	80417393 (2693, 2694)	Novel Protein sim. GBank gi 4504379 ref NP_003658.1 pHG38 - orphan G protein- coupled receptor HG38	Contains protein domain (PF00360) - glycoprotein Leucine Rich Repeat			264908, 264908, 264908, 265008, 264910, 265011, 265017, 264764, 264766, 264767, 264769, 264631, 264634, 264638, 264567, 264486
1348	87352335 (2695, 2696)	Novel Protein sim. GBank gi 3359720 dbj BAA321001  - (AB010999) peptidylarginine deiminase type IV [Rattus norvegicus]			UNCLASSIFIED	264488, 264489, 264508, 264509, 264510, 264511, 264512, 264591, 264592, 264601, 264684, 264685, 264769, 264532, 264534, 264555, 264556, 264557, 264558, 22279002, 264488

1349	91225548 (2697, 2698)	Novel Protein sim. GBank gij2144101 pir  55210 - tricarboxylate carrier - rat (fragment)		UNCLASSIFIED	52646842, 35695286, 22278996, 22278998, 22278999, 264259, 29331822, 28331824, 35696052, 284509, 284905, 284906, 284907, 284908, 284909, 264511, 265006, 264512, 284910, 60170831, 284591, 60433436, 284757, 21906754, 265017, 265018, 264605, 264760, 264762, 264288, 264768, 264689, 21906765, 21906766, 21906767, 21906768, 55811857, 35695917, 265020, 284534, 284691, 264692, 33657023, 264693, 33657349, 18108374, 18108376, 35696423, 60170394, 22279000, 22279002, 284563, 284564
1350	87093138 (2699, 2700)			UNCLASSIFIED	52646842, 264259, 29331825, 284908, 264511, 264604, 264288, 21906769, 265020, 33657182, 33657349, 18108374, 35695855, 284553, 264558, 18108385, 22279002, 284486
1351	87361327 (2701, 2702)	Novel Protein sim. GBank gij4887239 gb AAD32246.1  - (AF084594) BAW protein [Fugu rubripes]		UNCLASSIFIED	284908, 284907, 284638
1352	80076386 (2703, 2704)			UNCLASSIFIED	284693, 263981
1353	95345417 (2705, 2706)	Novel Protein sim. GBank gij2144101 pir  55210 - tricarboxylate carrier - rat (fragment)		UNCLASSIFIED	35698286, 60424269, 284603, 284509, 284906, 284907, 284908, 284909, 284511, 264512, 284910, 264756, 284598, 55811386, 265011, 284605, 55811150, 284782, 284784, 264766, 52644229, 56181562, 35695917, 265022, 33657023, 284693, 35695763, 60431528, 284828, 283978, 35696423, 35695855, 284630, 284634, 284635, 284636, 284637, 284638, 284639, 18108385, 284593, 284584, 284586
1354	95350845 (2707, 2708)	Novel Protein sim. GBank gij4688108 gb AAD27763.1 AF07703 - (AF07703) hypothetical 43.2 kDa protein [Homo sapiens]		UNCLASSIFIED	22278995, 22278999, 29331826, 284906, 285008, 33657402, 21906754, 265011, 87188559, 284684, 284388, 264789, 284689, 21906765, 21906768, 52644150, 33657023, 264692, 284693, 18108374, 83373044, 87188518, 22279000
1355	88260186 (2709, 2710)	Novel Protein sim. GBank gij146919 db BAA09487  - (D50928) The KIAA0138 gene product is novel. [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	dna_ma_bind	22278998, 22278997, 264259, 66714117, 284511, 21906754, 265010, 264769, 284689, 21906765, 21906768, 21906769, 264532, 27486262, 264629, 284638, 284556, 284638, 284639, 284482, 264484

1356	85313991 (2711, 2712)	Novel Protein sim. GBank gij113865 (U40342) - ninein [Mus musculus]		struct	18108397, 22276995, 22276996, 22276998, 264094, 29331828, 264905, 265006, 265007, 265008, 265010, 265017, 265018, 265019, 264764, 18108354, 264689, 21906785, 265022, 18108364, 35686423, 83373044, 18108387
1357	88260268 (2713, 2714)	Novel Protein sim. GBank gij897693[emb]CAA90330] - (Z50026) phosphatidylcholine transfer protein [Bos taurus]	Contains protein domain (PF01852) - START domain		264259, 29331822, 29331825, 264510, 87168559, 265018, 264448, 264288, 21906785, 21906786, 21906788, 265021, 284893, 18108378
1358	38719455 (2715, 2716)	Novel Protein sim. GBank gij556219 (L36831) - transcription regulator [Mus musculus]			264757
1359	87771643 (2717, 2718)			UNCLASSIFIED	264907, 264909, 264510, 264511, 264512, 18108351, 264764, 264534, 33657023, 18108374, 264634, 264635, 264638, 264639, 18108385, 264488, 264567
1360	87738272 (2719, 2720)	Novel Protein sim. GBank gij2588282[emb]CAA75812] - (Y15417) acetaldehyde-CoA ligase [Coprinus cinereus]		synthase	60432289, 264605
1361	87593527 (2721, 2722)	Novel Protein sim. GBank gij5589443[dbj]BAA830005.1] - (AB028976) KIAA1053 protein [Homo sapiens]	Contains protein domain (PF00538) - UNCLASSIFIED SAM domain (Sterile alpha motif)		35696286, 22276997, 22276999, 264259, 29331826, 264508, 264509, 264905, 264907, 264908, 265007, 265009, 33109954, 21906754, 87168474, 265011, 264761, 264683, 264288, 264766, 264769, 264689, 21906768, 265020, 265021, 33657023, 55811578, 35696423, 264634, 60432113, 22279002, 264482, 264486
1362	85287961 (2723, 2724)	Novel Protein sim. GBank gij5689411[dbj]BAA82888.1] - (AB028960) KIAA1037 protein [Homo sapiens]	Contains protein domain (PF00400) - eph WD domain, G-beta repeat		56182575, 56181606, 60432049, 264259, 29331822, 56182181, 29331827, 35696052, 29331828, 264905, 264906, 264908, 264595, 55812038, 85658542, 55811150, 264681, 264288, 264369, 56181582, 60431528, 55810764, 35696423, 60431850, 284559
1363	85758476 (2725, 2726)	Novel Protein sim. GBank gij1130484 (U35776) - ADP-ribosylation factor 1-directed GTPase activating protein [Rattus norvegicus]	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	UNCLASSIFIED	264488, 29331826, 264907, 264687, 264689, 264693
1364	88179488 (2727, 2728)				60432289, 60433356, 60433438, 87168559, 264603, 18108351, 21906786, 35696423, 60432113
1365	83003108 (2729, 2730)	Novel Protein sim. GBank gij4589562[dbj]BAA76803.1] - (AB023176) KIAA0959 protein [Homo sapiens]		oncogene	264766
1366	87003262 (2731, 2732)	Novel Protein sim. GBank gij1084944[pir]ISG4495 - hypothetical protein YPR021c - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00153) - transport Mitochondrial carrier proteins		265007

1367	87721210 (2733, 2734)	Novel Protein sim. GBank gij4894088[emb]CAB43240.1] - (AL050019) hypothetical protein [Homo sapiens]	Contains protein domain (PF01342) - SAND domain	UNCLASSIFIED	264488, 52646842, 52646365, 22278885, 56994075, 35698288, 22278998, 22278998, 22278998, 264259, 26331824, 66714117, 29331825, 60432289, 35698052, 264905, 264907, 264908, 264908, 264510, 265006, 265007, 264512, 264910, 265009, 60170831, 33657402, 55812038, 21908754, 265011, 87188359, 265017, 265019, 18108351, 264448, 264682, 264683, 264288, 264369, 264688, 264767, 264889, 21908755, 21908766, 21908768, 21908789, 55811857, 265020, 265021, 265022, 264534, 60170615, 264690, 264691, 18108352, 33657023, 33657109, 33657349, 264628, 18108370, 18108374, 18108376, 55811576, 35698423, 35698555, 264635, 264555, 264637, 264558, 52644332, 80170394, 264558, 18108381, 18108385, 59528488, 22279000, 264563, 264567
1368	94320078 (2735, 2736)	Novel Protein sim. GBank gij464561[sp]P35289[Rb15_RAT] RAS-RELATED PROTEIN RAB-15	Contains protein domain (PF00071) - Ras family	oncogene	264259, 29331822, 29331828, 60432288, 29331827, 35696052, 264508, 264805, 264908, 264908, 264909, 264510, 265007, 264910, 60433438, 264758, 89658542, 87168559, 264800, 264801, 264760, 264764, 264765, 264768, 52644229, 264689, 35695917, 265020, 265021, 294631, 284632, 264634, 264637, 52644332, 264558, 264839, 83373044, 264563, 264566, 264486, 264567
1369	86634033 (2737, 2738)	Novel Protein sim. GBank gij2062702 (U90550) - bulkyophilin [Homo sapiens]		UNCLASSIFIED	265008, 60432229, 60433356, 33657084, 21906764, 21906768, 264555, 264638, 264559, 264567
1370	95316810 (2739, 2740)	Novel Protein sim. GBank gij5031823[ref]NP_005823.1[PKCNM - potassium large conductance calcium-activated channel, subfamily M, beta member 2]		potassium_channel	22278996, 264259, 29331822, 29331824, 29331828, 29331827, 35696052, 264508, 264509, 264905, 264908, 264907, 264908, 264909, 264510, 264511, 264512, 264758, 265011, 265019, 264764, 264766, 264768, 21906767, 35695917, 18108362, 35696423, 264632, 264635, 264636, 264555, 264638, 264558, 264639, 18108385, 65274727, 264404, 264583, 264566, 264486
1371	95336512 (2741, 2742)	Novel Protein sim. GBank gij5032203[ref]NP_005714.1[PTSPA - tetraspan 5]	Contains protein domain (PF00335) - 4 transmembrane segments integral membrane proteins	glycoprotein	22278996, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 264805, 265009, 29331830, 66712502, 265008, 265009, 264758, 33657084, 85658542, 265010, 265018, 265019, 264782, 264448, 35695917, 33657109, 33657182, 33657349, 35695855, 264558, 22279002, 264563

1372	80248517 (2743, 2744)	Novel Protein sim. GBank gll840708[dbj BAA09334] - (D50685) trans-sialidase [Trypanosoma cruzi]		collagen	263978	264769, 21908765, 21908787, 22278999, 264691, 264910, 55812038, 265010, 264681, 264684
1373	80499421 (2745, 2746)			UNCLASSIFIED		
1374	95087038 (2747, 2748)	Novel Protein sim. GBank gll111878[pir JC1241] - beta-interferon-induced protein - rat		interferon		264907, 264510, 265007, 264512, 265008, 60432228, 264689, 65274781, 264555, 264556, 264557, 83373044, 60432113
1375	84235942 (2749, 2750)	Novel Protein sim. GBank gll5649176[gb AA003500.2] - (AF051155) G beta-like protein GBL [Rattus norvegicus]	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase		52644507, 52645196, 52646842, 52646365, 56182575, 58181686, 22278998, 56984075, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 52645080, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 33656970, 264905, 264509, 264906, 264907, 264908, 29331830, 264909, 265006, 264511, 265007, 265008, 265009, 33657402, 60433356, 52646317, 33108954, 33657084, 52644298, 87168474, 87168558, 264600, 265017, 265018, 265019, 55811150, 18108351, 264448, 264764, 264288, 264369, 264766, 52644229, 21908785, 21908786, 21908787, 21908789, 21908769, 55811957, 35695917, 265020, 265021, 52644150, 33657023, 264693, 65274820, 52845128, 33857109, 27486281, 33657348, 27486285, 35695783, 18108376, 55810764, 35696423, 35695855, 264630, 264631, 264634, 264636, 264555, 264638, 18108385, 87168518, 60432113, 22279000, 22279002, 264563, 264564, 264566, 264567
1376	87399050 (2751, 2752)	Novel Protein sim. GBank gll138350[sp P28869 VGLX_HSVEB - GLYCOPROTEIN X PRECURSOR		UNCLASSIFIED		284768, 284769, 35695917, 22278987, 284691, 284259, 29331822, 284693, 35695052, 284508, 284509, 284905, 264908, 284628, 284908, 284629, 18108372, 264909, 284510, 284511, 264512, 265008, 264630, 264631, 264910, 264632, 264634, 264635, 264636, 264591, 284592, 264637, 264638, 264558, 264639, 33657402, 264595, 18108385, 56526486, 265010, 265011, 264600, 284583, 284782, 284584, 264585, 264784, 284486, 284789
1377	86984242 (2753, 2754)	Novel Protein sim. GBank gll1663648 (U75321) - chromaffin granule ATPase II homolog [Mus musculus]	Contains protein domain (PF00122) - E1-E2 ATPase	ATPase-associated		29331824, 264591, 265019, 264686, 264788, 55811957, 264693, 22279002



1389	91256016 (2777, 2778)	Novel Protein sim. GBank gll5689387[dbj BAA82977.1] - (AB028948) KIAA1025 protein (Homo sapiens)	Contains protein domain (PF00641) - UNCLASSIFIED Zn-finger in Ran binding protein and others.	65274572, 22278999, 264259, 29331822, 29331824, 29331825, 60432289, 29331828, 29148499, 264906, 65712502, 55812038, 255017, 265018, 265019, 18108351, 264369, 21908765, 21908766, 21908767, 21908768, 265020, 265021, 264892, 33657023, 33657349, 18108370, 18108374, 55811576, 264555, 264556, 264557, 60170394, 83373044, 22279000, 264563, 264564, 52645156, 52646355, 264259, 52645080, 29331825, 29331826, 264906, 52644045, 265009, 33657084, 52644298, 87168474, 87168559, 265017, 265018, 264760, 264682, 264288, 264688, 264687, 56181562, 52644229, 21908765, 21908769, 35695917, 265020, 265021, 60170615, 52644150, 33657023, 27486261, 27486264, 65274791, 264631, 264555, 52644332, 87168518, 22279000, 264567
1390	94111916 (2779, 2780)	Novel Protein sim. GBank gll3702295 (AC005783) - R33083_1 [Homo sapiens]	peptidase	52645156, 52646355, 264259, 52645080, 29331825, 29331826, 264906, 52644045, 265009, 33657084, 52644298, 87168474, 87168559, 265017, 265018, 264760, 264682, 264288, 264688, 264687, 56181562, 52644229, 21908765, 21908769, 35695917, 265020, 265021, 60170615, 52644150, 33657023, 27486261, 27486264, 65274791, 264631, 264555, 52644332, 87168518, 22279000, 264567
1391	91227345 (2781, 2782)	Novel Protein sim. GBank gll1346910[sp P28650]PUA1, MOUSE - ADENYLOSUCCINATE SYNTHETASE, MUSCLE ISOZYME (IMP--ASPARTATE LIGASE)	Contains protein domain (PF00709) - Adenylosuccinate synthetase	52645156, 52646355, 264259, 52645080, 29331825, 29331826, 264906, 52644045, 265009, 33657084, 52644298, 87168474, 87168559, 265017, 265018, 264760, 264682, 264288, 264688, 264687, 56181562, 52644229, 21908765, 21908769, 35695917, 265020, 265021, 60170615, 52644150, 33657023, 27486261, 27486264, 65274791, 264631, 264555, 52644332, 87168518, 22279000, 264567
1392	94311097 (2783, 2784)	Novel Protein sim. GBank gll726286 (U22394) - mSin3A [Mus musculus]		52645156, 52646355, 264259, 52645080, 29331825, 29331826, 264906, 52644045, 265009, 33657084, 52644298, 87168474, 87168559, 265017, 265018, 264760, 264682, 264288, 264688, 264687, 56181562, 52644229, 21908765, 21908769, 35695917, 265020, 265021, 60170615, 52644150, 33657023, 27486261, 27486264, 65274791, 264631, 264555, 52644332, 87168518, 22279000, 264567
1393	80409472 (2785, 2786)		Contains protein domain (PF00560) - Leucine Rich Repeat	52645156, 52646355, 264259, 52645080, 29331825, 29331826, 264906, 52644045, 265009, 33657084, 52644298, 87168474, 87168559, 265017, 265018, 264760, 264682, 264288, 264688, 264687, 56181562, 52644229, 21908765, 21908769, 35695917, 265020, 265021, 60170615, 52644150, 33657023, 27486261, 27486264, 65274791, 264631, 264555, 52644332, 87168518, 22279000, 264567
1394	15028819 (2787, 2788)		UNCLASSIFIED	52645156, 52646355, 264259, 52645080, 29331825, 29331826, 264906, 52644045, 265009, 33657084, 52644298, 87168474, 87168559, 265017, 265018, 264760, 264682, 264288, 264688, 264687, 56181562, 52644229, 21908765, 21908769, 35695917, 265020, 265021, 60170615, 52644150, 33657023, 27486261, 27486264, 65274791, 264631, 264555, 52644332, 87168518, 22279000, 264567
1395	95351471 (2789, 2790)	Novel Protein sim. GBank gll2274845[dbj BAA21534] - (D88451) N-WASP [Rattus rattus]	UNCLASSIFIED	52645156, 52646355, 264259, 52645080, 29331825, 29331826, 264906, 52644045, 265009, 33657084, 52644298, 87168474, 87168559, 265017, 265018, 264760, 264682, 264288, 264688, 264687, 56181562, 52644229, 21908765, 21908769, 35695917, 265020, 265021, 60170615, 52644150, 33657023, 27486261, 27486264, 65274791, 264631, 264555, 52644332, 87168518, 22279000, 264567

1396	95363253 (2791, 2792)	Novel Protein sim. GBank gij2135904[pir][54810 - pHLEIF1 - human			22278997, 22278999, 264259, 29331825, 60432289, 29331828, 29146498, 29146499, 264907, 264908, 29331830, 264909, 265006, 265007, 265008, 265009, 60433356, 265010, 264602, 265017, 265018, 285018, 18108394, 52644229, 18108358, 21908767, 26148827, 21908768, 21906769, 29148629, 29148784, 265021, 265022, 18108368, 18108374, 56182323, 18108385, 264563, 264567, 35698286, 264907, 66712502, 264510, 35698917, 264892, 264693, 35698423, 264259, 29331822, 29331824, 29331825, 29331827, 35698052, 33656970, 87168474, 265018, 265019, 264682, 264766, 21908767, 265020, 33657023, 27486281, 55811576, 264632, 264639, 83373044, 87168518, 22279002
1397	87631317 (2793, 2794)			UNCLASSIFIED	
1398	91233667 (2795, 2796)	Novel Protein sim. GBank gij5420389[emb][CAB45680.1] - (AJ243460) proteophosphoglycan [Leishmania major]			284259, 29331822, 29331824, 29331825, 29331827, 35698052, 33656970, 87168474, 265018, 265019, 264682, 264766, 21908767, 265020, 33657023, 27486281, 55811576, 264632, 264639, 83373044, 87168518, 22279002
1399	87631078 (2797, 2798)	Novel Protein sim. GBank gij2496887[sp][Q09232]YQ22_CAEEL - HYPOTHETICAL 32.0 KD PROTEIN C09F5.2 IN CHROMOSOME III		UNCLASSIFIED	
1400	95419064 (2799, 2800)	Novel Protein sim. GBank gij283920[pir][S27939 - Iensin - chicken	Contains protein domain (PF00017) - Src homology domain 2	UNCLASSIFIED	56182576, 22278994, 22278997, 264259, 29331822, 29331825, 29331826, 29331827, 29331828, 264908, 56182435, 264112, 265009, 265011, 265017, 265018, 265019, 264760, 264762, 264765, 264288, 264685, 264687, 56181582, 264766, 21906766, 21906767, 55811857, 284891, 264692, 264628, 264629, 55811578, 264634, 264555, 264637, 264557, 264638, 18108381, 264556, 18108384, 60432113, 22278000
1401	91226379 (2801, 2802)	Novel Protein sim. GBank gij3256185[emb][CAA15485] - (AL008635) dJ510H16.1 [Homo sapiens]	Contains protein domain (PF00790) - VHS domain	UNCLASSIFIED	65274572, 60432289, 264909, 264759, 264768, 21908769, 22279002
1402	95361475 (2803, 2804)	Novel Protein sim. GBank gij1515427 (U57523) - nel homolog [Homo sapiens]	Contains protein domain (PF00008) - EGF-like domain	lgr	264905, 264907, 264908, 264909, 264112, 264693, 33657108, 264634
1403	94147933 (2805, 2806)	Novel Protein sim. GBank gij5262615[emb][CAB45747.1] - (AL080156) hypothetical protein [Homo sapiens]			65274572, 66712502, 265017, 264448, 264288, 21908765, 21908769, 264693, 55811576, 55274791, 60432113
1404	90935393 (2807, 2808)			UNCLASSIFIED	65274572, 22278998, 29331822, 29331828, 66712502, 265008, 60433438, 265017, 264693, 18108385



1405	95095068 (2809, 2810)	Novel Protein sim. GBank gij854065[emb]CAA58337] - (X83413) U88 [Human herpesvirus 8]			264488, 56994075, 35696286, 29331822, 29331824, 29331826, 29331828, 35696052, 264508, 264908, 264907, 264908, 264510, 264511, 264910, 33657402, 264594, 264758, 264600, 264604, 264762, 18108331, 264784, 33657023, 33957108, 264628, 264634, 83373044, 22279002, 264563, 264482, 264486, 264587, 264807, 264505
1406	87612368 (2811, 2812)	Novel Protein sim. GBank gij624076[g]AAC95425.1] - (U42580) contains Pro-rich Px motifs: SPKPP (20X), PEPPA (8X); similar to soybean pro-rich cell wall protein, corresponds to Swiss-Prot Accession Number P13993 [Paramecium bursaria Chlorella virus 1]	collagen		
1407	94129872 (2813, 2814)	Novel Protein sim. GBank gij2827886 (AF015037) - endoligopeptidase A related protein; EOPA related protein [Oryctolagus cuniculus]	UNCLASSIFIED		35696286, 22278998, 264094, 264259, 66714117, 29331826, 29331827, 29331828, 29146498, 264107, 264908, 265005, 265008, 264910, 60433438, 265011, 265017, 18108351, 264448, 264288, 264686, 21908785, 21908789, 264692, 33657109, 18108370, 264828, 263972, 18108374, 35696423, 55811576, 264631, 264557, 264558, 83373044, 18108385, 87168518, 60432113, 22279002
1408	85361477 (2815, 2816)	Novel Protein sim. GBank gij2554953 (AF030001) - unknown [Mus musculus]	oncogene	Contains protein domain (PF00008) - EGF-like domain	284488, 284489, 35696286, 264109, 264508, 264805, 264509, 264908, 264907, 264808, 264809, 265008, 265009, 264910, 33657402, 264757, 264758, 265011, 264501, 265017, 264760, 264762, 264683, 264685, 264766, 264687, 264689, 21906767, 265021, 264690, 264691, 33657023, 264692, 264693, 33657109, 264628, 264629, 35696423, 35695855, 264631, 264632, 264634, 264635, 264555, 264636, 264637, 264638, 56182323, 264639, 264583, 264584, 264585, 264586, 264587, 264593
1409	66644385 (2817, 2818)	Novel Protein sim. GBank gij2562165[db]BAA237141] - (AB007902) HH0712 cDNA clone for KIAA0442 has a 574-bp insertion at position 1474 of the sequence of KIAA0442 [Homo sapiens]			
1410	86612587 (2819, 2820)	Novel Protein sim. GBank gij2493790[sp]Q60994[ACR3.MOUSE - 30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN PRECURSOR (ACRP30) (ADIPOCYTE SPECIFIC PROTEIN ADIPOQ)]	complement	Contains protein domain (PF00386) - C1q domain	29331826, 264112, 264512, 265008, 265010, 264601, 264686, 264769, 21808767, 263974, 264631, 264586

1411	87818641 (2821, 2822)	Novel Protein sim. GBank gi 3123155 sp P01343 YM3M_CAEEL - HYPOTHETICAL 49.0 KD TRP-ASP REPEATS CONTAINING PROTEIN F35F8.5 IN CHROMOSOME I	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinasereceptor	22278995, 22278997, 22278998, 29331822, 60432289, 29331828, 264907, 265017, 265019, 264882, 21906767, 21906768, 21906769, 285020, 284890, 264691, 33657023, 33657109, 27486284, 264628, 263972, 264634, 264558, 18108385 264757
1412	84390918 (2823, 2824)	Novel Protein sim. GBank gi 387912 emb CAA94370  - (Z70310) predicted using GeneFinder. Similarity to Mouse ankyrin (PIR Acc. No. S37771); cDNA EST EMBL:TO1923 comes from this gene; cDNA EST EMBL:D32335 comes from this gene; cDNA EST EMBL:D32723 comes from this gene; cDNA EST E...	Contains protein domain (PF00023) - Ank repeat	UNCLASSIFIED homeobox	58994075, 29331822, 35696052, 29331828, 29331830, 264909, 52644045, 264510, 52644296, 85656542, 87188474, 265017, 265018, 264681, 264687, 21906768, 35695917, 265020, 52644150, 264692, 263967, 27486284, 35695763, 264638, 18108387, 264566
1414	94575860 (2827, 2828)	Novel Protein sim. GBank gi 3252981 (AF068921) - Ras- binding protein SUR-8 (Mus musculus)	Contains protein domain (PF00560) - Leucine Rich Repeat	UNCLASSIFIED	264682, 264683, 265022, 264636
1415	94326948 (2829, 2830)	Novel Protein sim. GBank gi 1871187 (U90439) - unknown protein (Arabidopsis thaliana)			52646365, 56182575, 22278994, 22278995, 58994075, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 29146498, 66712502, 29331830, 52644045, 264113, 264511, 33657402, 264757, 21906754, 55811386, 265017, 265018, 265019, 264761, 264683, 264369, 264288, 264686, 264689, 21906766, 21906767, 29148627, 21906769, 55811937, 265020, 265021, 264690, 33657023, 65274620, 52645129, 27486262, 27486284, 60431528, 264629, 35695855, 56182323, 264559, 60432113, 264404, 22279002, 264482

1416	84325977 (2831, 2832)	Novel Protein sim. GBank gll5106557/gb/AAD39749.1 (AF123052) MLL sepin-like fusion protein [Homo sapiens]	Contains protein domain (PF00735) - struct Cell division protein	18108392, 18108394, 18108397, 18108398, 22278995, 22278996, 35698286, 22278997, 22278998, 22278999, 29331822, 35696052, 29331828, 29146498, 264905, 264907, 264908, 264828, 264909, 264113, 265006, 265007, 265008, 265009, 60170831, 264595, 18108348, 21906754, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 264762, 18108351, 264681, 264763, 264682, 264683, 284766, 42644229, 264688, 264689, 21906765, 21906766, 21906767, 29146827, 21906768, 55811957, 29146829, 265020, 52644150, 18108361, 33657023, 18108362, 18108368, 264628, 18108370, 264629, 18108374, 18108376, 55811576, 65274791, 264634, 264636, 58182323, 18108381, 60170394, 18108385, 56526486, 87168518, 22279000 264107, 264448
1417	87826663 (2833, 2834)	Novel Protein sim. GBank gll4958935/dbjBAA78095.1 - (AB027570) suppressor of potassium transport defect 3 [Rattus norvegicus]	ATPase_associated	
1418	87594276 (2835, 2836)		UNCLASSIFIED	264259, 264808, 265010, 52844229, 21906764, 21906768, 264690, 264639, 18108388 264288
1419	87757168 (2837, 2838)	Novel Protein sim. GBank gll2072294 (U95097) - mitotic phosphoprotein 43 [Xenopus laevis]	struct	264259, 60432289, 265006, 87168474, 264288
1420	87298628 (2839, 2840)	Novel Protein sim. GBank gll5174421/reljNP_006023.1 (pCPNE - copine VI [neuronal] (Z69635) Similarity to Yeast uridine kinase (SW:URK1_YEAST); cDNA EST EMBL:Z14695 comes from this gene; cDNA EST CEMSE17F comes from this gene; cDNA EST EMBL:D67355 comes from this gene; cDNA EST yk209h1.5 comes from this gene...	ATPase_associated	29331824, 265007, 264563
1421	94746986 (2841, 2842)		kinase	18108358, 18108388, 18108397, 21906766, 18108398, 21906767, 58182575, 21906768, 21906769, 56181686, 55811957, 35695917, 35696286, 22278996, 22278997, 22278998, 22278999, 265021, 265022, 60170615, 264259, 33657023, 29331822, 56182181, 29331824, 66714117, 29331825, 33657109, 29331826, 27488281, 29331828, 35696052, 33657349, 264905, 264509, 20281149, 18108370, 264907, 60431528, 66712502, 263972, 55811576, 35696423, 35695855, 264512, 265007, 60431850, 60432229, 60431735, 58182323, 264558, 60170394, 83373044, 55812038, 264758, 18108385, 21906754, 55811386, 87168518, 87168559, 60432113, 265017, 265018, 265019, 22276002, 55811150, 264563, 264682, 264763, 264448, 264566, 264486, 18108391

1422	88178777 (2843, 2844)	Novel Protein sim. GBank gil4505939[re]NP_000928.1[pPOLR - polymerase (RNA) II (DNA directed) polypeptide A (220kD)]			mapolymerase	5694075, 35698286, 87168359, 55811957, 55811576, 284555, 284557, 87168518
1423	86997762 (2845, 2846)				UNCLASSIFIED	264686, 264489, 264692, 264594, 264603, 285018, 284908
1424	95201610 (2847, 2848)	Novel Protein sim. GBank gil437181 (U02289) - GTPase- activating protein [Caenorhabditis elegans]	Contains protein domain (PF00620) - RhoGAP domain		struct	29331822, 29331825, 29331827, 29146498, 264905, 264908, 264908, 264908, 285007, 284810, 285009, 33108854, 285010, 87168559, 285018, 284786, 264887, 21906765, 21906768, 21906767, 21906768, 29146627, 55811957, 29148529, 265021, 264691, 264692, 56526486, 22279002, 264563
1425	21662314 (2849, 2850)	Novel Protein sim. GBank gil100798[pr][S14859 - proline- rich protein - wheat]			UNCLASSIFIED	265007, 264558
1426	94322115 (2851, 2852)	Novel Protein sim. GBank gil2078441 (U56954) - weak similarity to S. cerevisiae intracellular protein transport protein US11 (SP:P25386) [Caenorhabditis elegans]			UNCLASSIFIED	264488, 60424179, 35699286, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 29331827, 56182435, 264910, 60433356, 60433438, 21906754, 265018, 264288, 21906765, 21906766, 21906767, 21906769, 265020, 265022, 33657109, 18108370, 18108376, 264558, 83373044, 18108385, 56526486, 22279002, 264482
1427	91227510 (2853, 2854)	Novel Protein sim. GBank gil5616074[gb]AAD45616.1[AF06194 - (AF06194.3) probable- derived STE20-like kinase PSK (Homo sapiens)]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain		kinase	22278994, 56994075, 22278997, 29331828, 29331830, 264828, 265006, 265007, 265008, 265009, 264683, 264288, 18108354, 21906785, 21906788, 29148629, 33657023, 18108374, 35699555, 83373044, 22278002, 264584
1428	94323008 (2855, 2856)	Novel Protein sim. GBank gil138350[sp]P28968[VLX_HSVB - GLYCOPROTEIN X PRECURSOR]			glycoprotein	56181686, 264259, 284907, 265007, 265009, 264595, 265010, 264686, 65274620, 264629, 65274791, 22279002, 264566
1429	87888689 (2857, 2858)				UNCLASSIFIED	264112, 264595, 265017, 265019, 21906765, 263977, 284555
1430	94735021 (2859, 2860)	Novel Protein sim. GBank gil1181619[db][BAA11565] - (D82384) a variant of TSC-22 [Gallus gallus]				264094, 29331824, 264591, 264593, 265018, 264681, 21906785, 21906787, 65274620, 55811576, 264639, 87168518, 22279002
1431	80429081 (2861, 2862)	Novel Protein sim. GBank gil5420389[emb][CAB46680.1] - (AJ243460) proteophosphoglycan [Leishmania major]			UNCLASSIFIED	264768, 265020, 284691, 264693, 264638
1432	87463004 (2863, 2864)	Novel Protein sim. GBank gil414797 (L18966) - pyruvate dehydrogenase phosphatase [Bos taurus]			phosphatase	18108394, 29146498, 265007, 60433438, 264763, 29148629, 263969
1433	87605403 (2865, 2866)	Novel Protein sim. GBank gil2460316 (AF022147) - uterus- ovary specific putative transmembrane protein [Rattus norvegicus]	Contains protein domain (PF00100) - Zona pellucida-like domain		UNCLASSIFIED	284259, 264510, 284591, 264603, 264585
1434	85713730 (2867, 2868)				UNCLASSIFIED	264682, 264691

1435	94708213 (2869, 2870)	Novel Protein sim. GBank gll3970850(db)[BAA34789.1] - (AB015330) HRIHF2007 (Homo sapiens)		transcript factor	22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 66714117, 29331828, 33856870, 264508, 264905, 66712502, 29331830, 264909, 265007, 265008, 264910, 265009, 60433356, 60433438, 264596, 21908754, 265010, 265017, 265018, 265019, 18108351, 264762, 264448, 264288, 264769, 21906767, 21906768, 21906769, 265020, 265021, 265022, 264690, 264691, 33657109, 264628, 18108374, 18108376, 55811576, 264636, 60170394, 56182323, 264559, 83373044, 87168518, 60432113, 22279000, 22279002, 264563, 264482, 264565
1436	86535024 (2871, 2872)	Novel Protein sim. GBank gll3183977(emb)[CAA39515] - (X55044) protein H19C [Mus musculus]		UNCLASSIFIED	22278997, 66714117, 29331826, 264907, 56182435, 265009, 18108351, 264692, 264693
1437	87631082 (2873, 2874)	Novel Protein sim. GBank gll2468687(sp)[Q09232](YQ22_CAEEL - HYPOTHETICAL 32.0 KD PROTEIN C09F5.2 IN CHROMOSOME III)		UNCLASSIFIED	264488, 264508, 264908, 264909, 264757, 264600, 264601, 264605, 264768, 264769, 264689, 18108394, 65274572, 56182575, 264489, 18108394, 65274572, 56182575
1438	85544280 (2875, 2876)	Novel Protein sim. GBank gll1505906 (AD000092) - hypothetical human serine-threonine protein kinase R31240_1 [Homo sapiens]		kinase	22278994, 22278996, 35666286, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331824, 60432269, 29331826, 29331827, 29331828, 35686052, 33856870, 264107, 264508, 264509, 264907, 66712502, 29331830, 56182435, 264511, 265008, 265007, 265009, 60432229, 6043438, 264595, 55812038, 55811386, 265011, 265017, 265018, 265019, 18108351, 264448, 18108354, 264288, 18108355, 264767, 21908765, 21908766, 21906767, 21906768, 21908769, 55811857, 35695917, 265020, 265021, 265022, 33657109, 18108370, 18108374, 55810764, 35695855, 264634, 264636, 56182323, 83373044, 18108387, 87168518, 60432113, 22279000, 264486
1439	91231864 (2877, 2878)	Novel Protein sim. GBank gll3876289(emb)[CAA94892] - (Z71180) similar to BP TUKUNITZ inhibitor domain; cDNA EST EMBL:D68293 comes from this gene; cDNA EST yk448h4.5 comes from this gene; cDNA EST yk24966.5 comes from this gene; cDNA EST yk448h4.3 comes from this gene [Caenorhabditis]		calyptin	22278994, 22278996, 35666286, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331824, 60432269, 29331826, 29331827, 29331828, 35686052, 33856870, 264107, 264508, 264509, 264907, 66712502, 29331830, 56182435, 264511, 265008, 265007, 265009, 60432229, 6043438, 264595, 55812038, 55811386, 265011, 265017, 265018, 265019, 18108351, 264448, 18108354, 264288, 18108355, 264767, 21908765, 21908766, 21906767, 21906768, 21908769, 55811857, 35695917, 265020, 265021, 265022, 33657109, 18108370, 18108374, 55810764, 35695855, 264634, 264636, 56182323, 83373044, 18108387, 87168518, 60432113, 22279000, 264486
1440	87423643 (2879, 2880)	Novel Protein sim. GBank gll2662165(db)[BAA23714] - (AB007802) HH0712 cDNA clone for KIAA0442 has a 574-bp insertion at position 1474 of the sequence of KIAA0442. [Homo sapiens]		UNCLASSIFIED	264887, 264258, 264908, 264907, 264908, 264909, 265008, 265010, 265017, 265018, 265019, 18108351, 264369, 265020, 33657023, 33657109, 60431528, 55811578, 264635

1441	95317662 (2881, 2882)	Novel Protein sim. GBank gij4493959[emb]CAB11123.2] - (Z98551) predicted using hexExon, MAL3P6.28 (PFC0845c). Hypothetical protein, len: 167 aa: Similarity to model organism hypothetical proteins (C.elegans, D.melanogaster, S.cerevisiae & S.pombe). C.elegans protein ZK287.5 (Tr....	Contains protein domain (PF00646) - helicase F-box domain.	18108392, 264488, 263994, 264489, 56182375, 22278994, 22278995, 56994075, 35696286, 22278997, 22278998, 60432049, 264259, 29331822, 29331824, 56182181, 68714117, 29331826, 29331827, 29331828, 35696052, 29146498, 264508, 284905, 264509, 264908, 264907, 264908, 264909, 58182435, 264510, 264511, 264512, 264910, 264592, 264593, 33657402, 60433438, 264595, 264758, 21906754, 85638542, 87168474, 265010, 87168558, 264600, 264602, 265017, 284604, 285018, 284605, 285019, 264760, 264761, 264762, 264881, 264448, 264764, 264683, 264288, 264768, 264769, 264769, 52844229, 264889, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35895917, 265020, 265021, 60170615, 52844150, 264891, 264692, 33657023, 65274620, 33857109, 35695763, 264628, 18108370, 264629, 18108374, 55811576, 35696423, 65274781, 35695855, 264631, 264634, 264635, 264638, 264637, 264638, 56182323, 264558, 60170394, 264639, 264559, 83373044, 18108385, 18108388, 56526486, 22279000, 22279002, 264563, 264483, 264564, 264566, 264567, 264489
1442	83367491 (2883, 2884)	Novel Protein sim. GBank gij5103027[dj]BAA78765.1] - (AB023419) mSox7 [Mus musculus]	transcription factor	264906, 265007, 264693, 264558
1443	87109935 (2885, 2886)	Novel Protein sim. GBank gij4887229[gb]AAD32244.1[AF150755] - (AF150755) microtubule-actin crosslinking factor [Mus musculus]	Contains protein domain (PF00435) - spectrin repeat	52845080, 264691, 264628, 264555
1444	87620478 (2887, 2888)	Novel Protein sim. GBank gij3874447[emb]CAB02772] - (Z81039) predicted using Genefinder: cDNA EST EMBL: T01209 comes from this gene: cDNA EST yk278a11.3 comes from this gene: cDNA EST yk308a9.3 comes from this gene: cDNA EST yk308a9.5 com...	UNCLASSIFIED	264259, 29331822, 29331824, 66714117, 29331828, 264288, 35695917, 33657023, 264635, 80170394
1445	94990470 (2889, 2890)	Novel Protein sim. GBank gij2959886[emb]CAA11022] - (AJ222868) L-peritaxin [Mus musculus]	UNCLASSIFIED	264369
1446	85079268 (2891, 2892)	Novel Protein sim. GBank gij5081610[gb]AAD39484.1[AF135440] - (AF135440) huntingtin yeast partner C [Mus musculus]	Contains protein domain (PF01846) - FF domain	264369
1447	86945392 (2893, 2894)			18108396, 35696286, 22278997, 68714117, 29331828, 265009, 264758, 285018, 264288, 21906766, 21906767, 264692, 264634, 264566
1448	94990477 (2895, 2896)	Novel Protein sim. GBank gij3880411 (AC004551) - putative proline-rich protein [Arabidopsis thaliana]	Contains protein domain (PF00439) - Bromodomain	29331827, 264509, 264909, 265008, 264595, 18108357, 18108385, 264566, 264486

1448	8760859 (2897, 2898)				UNCLASSIFIED	66714117, 284908, 284908, 284591, 284601, 284784, 284632
1450	8745896 (2899, 2900)	Novel Protein sim. GBank gji1707074 (U80450) - M01E11.2 [Caenorhabditis elegans]			UNCLASSIFIED	35686288, 35696052, 265008, 265009, 60170831, 33109954, 264683, 264689, 35686423, 35695855, 56626486
1451	87767970 (2901, 2902)	Novel Protein sim. GBank gji160304[emb]CAA10600] - (AJ132192) HS1 binding protein 3 [Mus musculus]			UNCLASSIFIED	29331826, 284683, 284693, 263978, 284630
1452	85602899 (2903, 2904)	Novel Protein sim. GBank gji2832906[db]BAA24508.1] - (D89340) dipeptidyl peptidase III [Rattus norvegicus]			peptidase	284681, 33857023, 284629
1453	88130434 (2905, 2906)	Novel Protein sim. GBank gji728831[sp]P39188[ALU1_HUMAN - III] ALU SUBFAMILY J WARNING ENTRY !!!			kinase	264510, 264768
1454	11204696 (2907, 2908)					264556
1455	87757896 (2909, 2910)				UNCLASSIFIED	29331822, 66714117, 29331825, 284905, 29331830, 265006, 265008, 265009, 265011, 265019, 18108351, 21906768, 33657109, 18108376, 264632, 56182323, 87168518
1456	86320218 (2911, 2912)	Novel Protein sim. GBank gji729230[sp]P41004[CUT3, SCHPO - CHROMOSOME SEGREGATION PROTEIN CUT3			transport	22278995, 22278996, 22278997, 22278998, 22278999, 29331827, 264107, 265017, 21906765, 21906766, 21906767, 21906769, 29148629, 18108370, 22278000
1457	80076900 (2913, 2914)				UNCLASSIFIED	264107, 264568
1458	87800460 (2915, 2916)	Novel Protein sim. GBank gji2246532 (U80872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]			transport	56182575, 22278999, 60432049, 264259, 29331826, 29331827, 29331828, 264102, 264107, 264110, 265009, 60432229, 265019, 265020, 263972, 263976, 264635, 22279002, 264568
1459	95360920 (2917, 2918)	Novel Protein sim. GBank gji5524667[gb]AAD4333.1[AF15935 - (AF159356) Munc13-4 protein [Rattus norvegicus]	Contains protein domain (PF00168) - C2 domain		kinase	22278997, 264259, 29331824, 29331826, 29331827, 29331828, 265017, 265018, 264760, 264682, 264448, 264288, 264768, 265021, 264692, 33657023, 33657109, 35695855, 264566
1460	95346602 (2919, 2920)				UNCLASSIFIED	29331822, 264591, 55811957, 264691, 264693, 65274620
1461	94741513 (2921, 2922)	Novel Protein sim. GBank gji1707274 (U80931) - strong similarity to class-III of pyridoxal-phosphate dependent aminotransferases [Caenorhabditis elegans]	Contains protein domain (PF00202) - Aminotransferases class-III pyridoxal phosphate		gaba	22278997, 29331822, 35686052, 265009, 264756, 265017, 265018, 265019, 264760, 264399, 264687, 21906765, 21906768, 265022, 33657109, 27488261, 264555, 83373044
1462	87732018 (2923, 2924)				UNCLASSIFIED	264555, 264556
1463	88080605 (2925, 2926)	Novel Protein sim. GBank gji170466[emb]CAA66912] - (X98259) M-phase phosphoprotein 8 [Homo sapiens]	Contains protein domain (PF00385) - 'chromo' (CHR)romatin Organization Modifier domain		struct	50432048, 264259, 29146499, 264906, 264907, 264512, 265017, 264763, 264766, 18108370, 18108374, 264636, 18108385, 18108388

1464	87620482 (2827, 2926)	Novel Protein sim. GBank gij3874447[emb CAB02772] - (Z81039) predicted using GeneFINDER: cDNA EST EMBL: T01209 comes from this gene; cDNA EST yk278a11.3 comes from this gene; cDNA EST yk308a9.3 comes from this gene; cDNA EST yk308a9.5 comes from this gene; cDNA EST yk308a9.5 comes from this gene...		UNCLASSIFIED	264569, 22278995, 22278996, 22278997, 22278998, 29331822, 29331824, 29331825, 35698052, 20281100, 264905, 29331830, 264909, 265007, 33657402, 21906754, 265017, 265018, 264682, 264684, 264368, 264288, 264766, 21906769, 35695917, 264691, 21906767, 21906769, 35695917, 264691, 33657023, 264692, 35696423, 35695855, 264630, 264631, 264639, 264565
1465	87425192 (2926, 2930)	Novel Protein sim. GBank gij4589598[db BAA76821.1] - (AB023194) KIAA0977 protein [Homo sapiens]		glucoamylase	264488, 22278994, 56994075, 60432049, 264259, 56182181, 60432289, 29331827, 52844045, 264511, 265007, 265008, 264596, 55812038, 55811386, 264600, 264602, 265017, 265018, 264604, 265019, 18108351, 18108354, 56181562, 21906769, 265021, 33657023, 33657182, 55811576, 264557, 18108382, 60432113
1466	87606227 (2831, 2932)	Novel Protein sim. GBank gij2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		UNCLASSIFIED	284512, 285017, 284689, 264558
1467	87614328 (2933, 2934)	Novel Protein sim. GBank gij4507241[ra NP_003137.1]pSSRP - structure specific recognition protein 1 [Drosophila melanogaster]		struct	284683, 284636
1468	95342862 (2935, 2936)	Novel Protein sim. GBank gij4507241[ra NP_003137.1]pSSRP - structure specific recognition protein 1 [Drosophila melanogaster]			22278998, 264758, 265018, 265019, 21906769, 265020, 33657109, 22279002
1469	78236174 (2937, 2938)	Novel Protein sim. GBank gij1906596 (U81768) - kinesin-73 [Drosophila melanogaster]		struct	18108394, 18108397, 18108398, 35896052, 29146499, 265007, 265008, 265009, 265010, 265011, 18108354, 18108365, 18108368, 18108374, 18108381, 18108382, 18108384, 18108388
1470	94890482 (2939, 2940)	Novel Protein sim. GBank gij5849170[gb AAD43131.2]AF15909 - (AF159092) syld708613 protein [Homo sapiens]		UNCLASSIFIED	18108394, 18108398, 56182575, 264259, 29331822, 29331824, 29331825, 60432289, 264907, 264909, 265007, 264910, 265009, 264591, 60432229, 60433358, 264595, 60433438, 264758, 33108954, 265010, 265011, 265018, 264760, 264448, 264764, 264288, 264369, 18108357, 264769, 18108358, 21906767, 21906769, 55811957, 265021, 18108361, 264691, 18108362, 18108365, 18108368, 264628, 18108379, 264637, 264597, 18108381, 56182323, 18108382, 83373044, 18108384, 18108388, 87168518, 60432113, 264404, 22279002, 264482, 264567, 264487
1471	87826842 (2941, 2942)	Novel Protein sim. GBank gij3876146[emb CAB01750] - (Z78542) similar to Mitochondrial carrier proteins: cDNA EST EMBL: T01651 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	29331822, 29331824, 29331825, 264828, 264603, 264689, 264693, 18108374, 55811576



1472	87756616 (2943, 2944)	Novel Protein sim. GBank gi 4680707 gb AA027743.1 AF13286 - (AF13286) CGI-34 protein [Homo sapiens]			UNCLASSIFIED	264905, 18106351, 21906765, 264486
1473	87751609 (2945, 2946)	Novel Protein sim. GBank gi 3568780 (AF042180) - testis-specific Y-encoded-like protein [Mus musculus]		Contains protein domain (PF00956) - Nucleosome assembly protein (NAP)	MHC	18106394, 22278995, 56994075, 22278999, 29331822, 29331824, 66714117, 29331825, 29331826, 35696052, 264906, 264907, 56182435, 265007, 264758, 265018, 265019, 264760, 264764, 264288, 264885, 264688, 264768, 21906769, 55811957, 265021, 264691, 264693, 264629, 55811578, 264634, 264636, 56182323, 22279002, 264566, 264486
1474	85800989 (2947, 2948)	Novel Protein sim. GBank gi 2494890 sp Q92176 CORO_BOVIN - CORONIN-LIKE PROTEIN P57		Contains protein domain (PF00400) - WD domain, G-beta repeat	struct	264488, 35695917, 35698288, 264682, 33657023, 264693, 33657109, 35698052, 264508, 264905, 264906, 264907, 264629, 264908, 264909, 35698423, 35695855, 264511, 264910, 264632, 264634, 264635, 264636, 264637, 264556, 264557, 264639, 264758, 60432113, 264604, 264605, 264565, 264566, 264764, 264486, 264685, 264765, 264681, 264682, 264288, 264566
1475	86871935 (2949, 2950)			Contains protein domain (PF00041) - Fibronectin type III domain	UNCLASSIFIED	
1476	87548855 (2951, 2952)	Novel Protein sim. GBank gi 4757732 ref NP_004664.1 pANGP - angiopoietin 3		Contains protein domain (PF00147) - Fibrinogen beta and gamma chains, C-terminal globular domain	glycoprotein	60424179, 55181686, 29331824, 60424269, 29331826, 35698052, 264508, 264905, 264906, 264907, 264908, 264909, 264512, 265007, 265008, 265009, 264910, 33657402, 264595, 264596, 55812038, 265011, 264601, 264762, 18106351, 264288, 264369, 264685, 264768, 264689, 55811857, 264691, 264692, 264693, 18106370, 60431528, 18108374, 35696423, 264634, 264635, 264636, 60431850, 264555, 264638, 264557, 264639, 18108382, 18108388, 60432113, 22279002, 264259, 264107, 264905, 265008, 265010, 265011, 264682, 264288, 265020, 265021, 263974
1477	87774279 (2953, 2954)	Novel Protein sim. GBank gi 2498308 sp Q60870 DPL_MOUSE - POLYPOSIS LOCUS PROTEIN 1 HOMOLOG (TB2 PROTEIN HOMOLOG) (GP106)			UNCLASSIFIED	
1478	11754412 (2955, 2956)					264686

1479	91640140 (2857, 2858)	Novel Protein sim. GBank gi5499741 gb AAD43978.1 AF152986 - (AF152986) chromatin-specific transcription elongation factor FACT 140 kDa subunit [Homo sapiens]		peptidase	56182575, 22278995, 22278996, 22278998, 22278999, 29331822, 29331824, 66714117, 264908, 264907, 58182435, 265006, 60170831, 33657402, 264758, 33109854, 21906754, 265017, 265018, 264448, 264288, 264767, 264887, 52644229, 21906764, 264689, 21906765, 21906768, 21906769, 265020, 265021, 60170815, 264891, 33657023, 33657109, 33657182, 27486281, 27486282, 33657348, 18108370, 60431528, 263976, 55811578, 264556, 264557, 60170394, 87168518, 264404, 22279000, 22279002, 264563, 264482
1480	94312412 (2959, 2960)	Novel Protein sim. GBank gi3550456 emb CAA08329.1  - (AJ005073) Alix [Mus musculus]		UNCLASSIFIED	18108394, 56274572, 56182575, 22278995, 35696286, 56994075, 22278996, 22278997, 22278998, 22278999, 284091, 264259, 35696052, 28146499, 264103, 264105, 284108, 264907, 52844045, 264112, 265007, 265008, 265009, 60433356, 60433438, 264596, 33109954, 33657084, 52644286, 87168474, 265010, 87168559, 265017, 265018, 265019, 264448, 264682, 264683, 264759, 21906765, 21906768, 21906767, 21906768, 21906789, 285020, 285021, 60170615, 52644150, 33657109, 33657182, 263972, 35695855, 264557, 263981, 83373044, 18108385, 87168518, 264566, 264088, 264486
1481	87021442 (2961, 2962)	Novel Protein sim. GBank gi4636807 gb AAD30566.1 AF14679 - (AF14679) PFT27 [Mus musculus]		MHC	265006, 265007, 265010, 18108374
1482	85320442 (2963, 2964)	Novel Protein sim. GBank gi4585372 gb AAD25403.1 AF12292 - (AF12292) Vmt inhibitory factor-1 [Mus musculus]	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	284608, 264910, 264758
1483	94115503 (2965, 2966)	Novel Protein sim. GBank gi535428 (U13736) - calmodulin- like protein [Pisum sativum]	Contains protein domain (PF00036) - EF hand	struct	264259, 29331822, 52645080, 29331825, 29331826, 33656970, 29331830, 265007, 55812038, 33109854, 285017, 264288, 21906768, 21906769, 264638, 18108380, 87168518, 22279000
1484	94131544 (2967, 2968)	Novel Protein sim. GBank gi1911774 bbs 180090 - (S83364) putative Rab5-interacting protein (clone L1-57) [human, HeLa cells, Peptide Partial, 122 aa] [Homo sapiens]		UNCLASSIFIED	264489, 35696286, 264259, 264107, 264909, 265008, 60433356, 33657402, 60433438, 264288, 21906765, 21906766, 29148627, 33657023, 27486262, 18108374, 35696423, 83373044, 60432113
1485	80194441 (2969, 2970)	Novel Protein sim. GBank gi5360129 gb AAD42883.1 AF15511 - (AF15511) NY-REN 62 antigen [Homo sapiens]	Contains protein domain (PF00225) - Kinesin motor domain	struct	264369, 265020, 18108374

1486	94125068 (2071, 2972)	Novel Protein sim. GBank gij4589518dbjBAA76780.1] - (AB023153) KIAA0936 protein [Homo sapiens]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	56182575, 22278999, 264906, 264907, 21906754, 87168474, 265017, 265019, 18108351, 264288, 265020, 264566, 21906754, 264486
1487	86452711 (2973, 2974)	Novel Protein sim. GBank gij5019275jembjCAB4431.1] - (AJ132751) xenobiotic/medium-chain fatty acid-CoA ligase [form XL-III] [Bos taurus]		synthase	
1488	87732026 (2975, 2976)	Novel Protein sim. GBank gij5712131gbjAAD47378.1] (AF120499) DEM1 protein [Homo sapiens]	Contains protein domain (PF01443) - Viral (Superfamily 1) RNA helicase	lgi	264686, 264769, 264689, 264682, 264683, 264509, 264906, 264907, 18108370, 264908, 264629, 264909, 264510, 265008, 264512, 265007, 265008, 265009, 264555, 264556, 264557, 264558, 264782, 264584, 264682, 21906767, 22278999, 265022, 264259,
1489	95104277 (2977, 2978)	Novel Protein sim. GBank gij24973035pQ62786jFPRP_RAT - PROSTAGLANDIN F2-ALPHA RECEPTOR REGULATORY PROTEIN PRECURSOR (PROSTAGLANDIN F2-ALPHA RECEPTOR ASSOCIATED PROTEIN)	Contains protein domain (PF00047) - Immunoglobulin domain	prostaglandin	264683, 29331824, 29331825, 29331826, 29331827, 29331828, 264103, 263972, 66712502, 35696423, 35695855, 265007, 265008, 265009, 83373044, 21906754, 56526486, 265017, 264563, 18108351, 264564, 264586, 264389, 264288
1490	87390127 (2979, 2980)			UNCLASSIFIED	56182575, 264259, 29331822, 29331824, 66714117, 29331827, 29331828, 264508, 264905, 66712502, 265007, 265008, 264594, 33657402, 55812038, 87168474, 265018, 18108351, 264369, 264288, 264769, 264689, 21906767, 21906768, 55811957, 60170615, 33657109, 35695855, 264635, 60170394, 56526486, 22279002, 264563
1491	83594305 (2981, 2982)	Novel Protein sim. GBank gij295671 (L11275) - selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III [Saccharomyces cerevisiae]		UNCLASSIFIED	265007, 264448, 18108372, 264568, 56182323
1492	85805363 (2983, 2984)	Novel Protein sim. GBank gij1656005 (U71205) - rit [Mus musculus]	Contains protein domain (PF00071) - Ras family	oncogene	22278997, 22278998, 29331822, 264907, 66712502

1493	91677215 (2985, 2986)	Novel Protein sim. GBank gij5689515[dbj]BA4304.1.1] - (AB029012) KIAA1089 protein [Homo sapiens]		UNCLASSIFIED	264488, 52846365, 65274572, 56182575, 22278994, 35696286, 56994075, 22278999, 60432049, 29331824, 28331828, 35696032, 264508, 264905, 264906, 52844045, 264909, 56182435, 265006, 265008, 265009, 60170831, 33657402, 55812038, 265010, 265011, 265017, 265018, 265019, 55811150, 264448, 264682, 264685, 264688, 52644229, 21906765, 21906768, 21906769, 21906769, 265020, 265021, 60170615, 52644150, 33657023, 18108384, 18108365, 33657109, 33657182, 27486281, 27486282, 27486284, 33657349, 27486285, 35693763, 18108370, 264629, 18108374, 52844332, 56182323, 87168518, 22279002, 264584, 264586, 264587
1494	87605285 (2987, 2988)	Novel Protein sim. GBank gij728832[sp]P39189[ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII]	Contains protein domain (PF01352) - Kinase KRAB box		264907, 265009, 264769, 18108370, 55811576, 264639, 264565, 264486
1495	87605267 (2989, 2990)	Novel Protein sim. GBank gij458958[dbj]BA4768.1.1] - (AB023189) KIAA0972 protein [Homo sapiens]	Contains protein domain (PF01352) - transcriptfactor KRAB box		22278997, 264259, 264906, 264907, 265009, 264594, 33657084, 265017, 264760, 264448, 33657108, 284630, 264634, 56528486, 264563, 264565, 264566, 264486, 264567
1496	87784322 (2991, 2992)	Novel Protein sim. GBank gij5420387[emb]CA46678.1] - (AJ243459) proteophosphoglycan [Leishmania major]	UNCLASSIFIED		35696286, 264906, 265019, 264693
1497	81695428 (2993, 2994)	Novel Protein sim. GBank gij3874825[emb]CA492591] - (Z68296) Similarly to Mouse A-RAF proto-oncogene serine/threonine-protein kinase (SW:KRAA_MOUSE); cDNA EST EMBL D27610 comes from this gene; cDNA EST EMBL T01018 comes from this gene; cDNA EST EMBL D33258 comes from this gene;....	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain		264910, 264758, 265011, 264764, 264288, 284690, 284634, 264635, 56526486
1498	80934938 (2995, 2996)	Novel Protein sim. GBank gij728838[sp]P39193[ALU6_HUMAN - IIII ALU SUBFAMILY SP WARNING ENTRY IIII]	oncogene		264488, 65274572, 28331822, 265017, 265018, 21906765, 29148827, 29148629, 18108374, 284637, 284638, 264567
1499	86451569 (2997, 2998)	Novel Protein sim. GBank gij2570198 (U34556) - microfilament sheath protein SHP3 [Litomosoides sigmodontis]	glucosylase		263978, 264568
1500	80499386 (2999, 3000)	Novel Protein sim. GBank gij2078483 (U43200) - antifreeze glycoprotein precursor [Boreogadus saida]	UNCLASSIFIED		22278998, 264769, 18108379, 284559
1501	85795297 (3001, 3002)	Novel Protein sim. GBank gij2078483 (U43200) - antifreeze glycoprotein precursor [Boreogadus saida]	UNCLASSIFIED		
1502	80206141 (3003, 3004)				264508, 264112, 284604, 284684, 52644150, 55811576, 284632, 284556, 264638, 56182323, 284563, 264486
1503	87012701 (3005, 3006)	Novel Protein sim. GBank gij3900855 (AC004874) - similar to N-acetylgalactosaminyltransferase; similar to Q07537 (PID:g1171989) [Homo sapiens]	Contains protein domain (PF00535) - transferase Glycosyl transferases		28331822, 265007, 264369

1504	79640031 (3007, 3008)		Contains protein domain (PF00023) - UNCLASSIFIED	264693	
1505	86102672 (3009, 3010)	Novel Protein sim. GBank gll4753775[emb]CABA1970.1] - (AJ132545) protein kinase [Homo sapiens]	Contains protein domain (PF00069) - UNCLASSIFIED	28331826, 35696052, 264509, 264906, 264907, 264908, 264909, 264511, 264910, 35695917, 265022, 33657109, 18108374, 264631, 264635, 264638, 264568	kinase
1506	94143218 (3011, 3012)	Novel Protein sim. GBank gll1304201[dbj]BAA06170] - (D29766) alternatively spliced product [Rattus norvegicus]	Contains protein domain (PF00018) - UNCLASSIFIED	55274572, 56182575, 56994075, 22278997, 22278998, 22278999, 264091, 264092, 60432049, 264259, 52645080, 29331822, 29331827, 264106, 29331830, 264908, 56182435, 264110, 264511, 264512, 55812038, 21908754, 87168559, 264600, 265017, 265018, 264681, 18108354, 264369, 264687, 264689, 21908765, 29148627, 21906768, 21906769, 29148629, 52644150, 33657023, 18108376, 65274791, 56182323, 264558, 264559, 18108385, 87168518, 60432113, 22279000, 264555	glycoprotein
1507	83736250 (3013, 3014)	Novel Protein sim. GBank gll5689513[dbj]BAA33040.1] - (AB025011) KIAA1088 protein [Homo sapiens]	SH3 domain	264639	helicase
1508	11618758 (3015, 3016)	Novel Protein sim. GBank gll5031975[ref]NP_005875.1[lp]PAK4 - protein kinase related to S. cerevisiae STE20, effector for Cdc42Hs	Contains protein domain (PF00069) - UNCLASSIFIED	264593	kinase
1509	87318451 (3017, 3018)	Novel Protein sim. GBank gll113161[sp]P28614[ACOR_ALCEU - ACETON CATABOLISM REGULATORY PROTEIN	Eukaryotic protein kinase domain	264259, 29331822, 29331824, 29331828, 264591, 33109954, 264563	
1510	95362643 (3019, 3020)	Novel Protein sim. GBank gll728631[sp]P3918[ALU1_HUMAN - IIII ALU SUBFAMILY J WARNING ENTRY IIII	UNCLASSIFIED	264259, 29331822, 265007, 18108374, 264556	
1511	88318073 (3021, 3022)	Novel Protein sim. GBank gll4559353[gb]AAD23014.1[AC006585] putative extragenic suppressor protein [Arabidopsis thaliana]	UNCLASSIFIED	264259, 29331822, 29331824, 29331827, 284509, 284907, 264510, 264511, 265007, 264512, 264008, 87168559, 264288, 265022, 33657023, 35695855, 264637, 264638, 284563	
1512	95346390 (3023, 3024)	Novel Protein sim. GBank gll4559353[gb]AAD23014.1[AC006585] putative extragenic suppressor protein [Arabidopsis thaliana]	Contains protein domain (PF01163) - UNCLASSIFIED	52645156, 18108386, 56994075, 60432289, 265006, 60433358, 60433438, 21908754, 87168474, 87168559, 265018, 264782, 264783, 264687, 21906765, 21906789, 27486262, 35695763, 18108374, 35696423, 264555, 18108385, 18108387, 18108388, 87168518, 264482	
1513	87436228 (3025, 3026)	Novel Protein sim. GBank gll1330394 (U58761) - C01F1.6 gene product [Caenorhabditis elegans]		35696052, 264905, 264906, 264907, 264908, 264909, 264910, 264591, 264766, 264689, 264692, 264629, 264636	

1514	95345392 (3027, 3028)	Novel Protein sim. GBank gi 4559353 gb AAD23014.1 AC00658 - (AC006585) putative extragenic suppressor protein [Arabidopsis thaliana]	Contains protein domain (PF01163) - RIO1/ZK632.3/MJ0444 family	UNCLASSIFIED	52644507, 52645156, 52646365, 52646842, 65274572, 22278994, 35698286, 56994075, 264259, 52645080, 29331822, 29331825, 35696052, 29331830, 52644045, 56182435, 265006, 60433358, 60433438, 55812038, 21906754, 52646317, 52644296, 87168474, 87168559, 284448, 52644229, 21906765, 21906766, 21906767, 21906768, 35695917, 265020, 52644150, 33657023, 52645128, 33657109, 33657182, 27486261, 27486262, 27486264, 27486265, 35695763, 18108376, 35698423, 35695855, 52644332, 18108385, 18108387, 87168518, 60432113 265020, 264639
1515	79163536 (3028, 3030)	Novel Protein sim. GBank gi 3679501 emb CA87795  - (Z47812) similar to ubiquitin carboxyl-terminal hydrolase; cDNA EST EMBL:D33365 comes from this gene; cDNA EST EMBL:D33965 comes from this gene; cDNA EST EMBL:D33822 comes from this gene; cDNA EST EMBL:D34547 comes from this ge...	ubiquitin		
1516	88073539 (3031, 3032)	Novel Protein sim. GBank gi 488015 (L27479) - X123 [Homo sapiens]	UNCLASSIFIED		265008, 56182323, 22279002
1517	87793925 (3033, 3034)	Novel Protein sim. GBank gi 3415134 (AF082024) - Phyo1 [Pimpinella brachycarpa]			264091, 18108370, 264404
1518	87350697 (3035, 3036)	Novel Protein sim. GBank gi 28838 sp P39195 ALU8_HUMAN - IIII ALU SUBFAMILY SX WARNING ENTRY IIII	im7		66714117, 264508, 264509, 264805, 264510, 264910, 264591, 264595, 264288, 264766, 264769, 18108374, 264638, 264639, 264486 264569, 264489, 60432049, 265009,
1519	94328689 (3037, 3038)	Novel Protein sim. GBank gi 526268 emb CA845771.1  - (AL080198) hypothetical protein [Homo sapiens]			33657402, 264598, 21908754, 265018, 264369, 21908785, 21908788, 21908789, 264691, 65274620, 33657182, 27488281, 18108374, 264557, 264639, 87168518, 22279002
1520	87582855 (3039, 3040)	Novel Protein sim. GBank gi 2662161 db BAA23712  - (AB007900) rH0452 cDNA clone for KIAA0440 has a 438- bp insertion at position 1711 of the sequence of KIAA0440. [Homo sapiens]	UNCLASSIFIED		18108392, 60432049, 264259, 29331824, 265007, 60433356, 265010, 21906768, 264638
1521	86970696 (3041, 3042)	Novel Protein sim. GBank gi 5052351 gb AAD38516.1 AF13542 - (AF135421) GDP- mannose pyrophosphorylase B [Homo sapiens]	Contains protein domain (PF00483) - Nucleotidyl transferase	synthase	18108394, 264259, 66714117, 265011, 264603, 265019, 18108364, 35698423, 264557, 264558, 18108388
1522	76980687 (3043, 3044)	Novel Protein sim. GBank gi 3776567 (AC005388) - Strong similarity to F2187.33 gi 2809264 from A. thaliana BAC gb AC002560. EST gb N65119 comes from this gene. [Arabidopsis thaliana]	UNCLASSIFIED		29331824, 265018, 265020, 265021
1523	91005151 (3045, 3046)		UNCLASSIFIED		65274572, 21906766, 264693
1524	80203723 (3047, 3048)				
1525	87799667 (3049, 3050)	Novel Protein sim. GBank gi 475904 ref NP_004283.1 pRIN1 - ras inhibitor	UNCLASSIFIED		264112, 21906754, 263974
			UNCLASSIFIED		264683, 264687, 264688, 264690, 264692, 264693

1526	95105344 (3051, 3052)	Novel Protein sim. GBank gi 72885050 p08640 AMYH_YEAST - GLUCOAMYLASE S1/S2 PRECURSOR (GLUCAN 1,4-ALPHA- GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE)		glycoprotein	35695286, 56182181, 60431735, 264595, 55812038, 264605, 264683, 21908785, 55811957, 265020, 65274791, 284535, 264556, 264557, 264558, 264559, 83373044
1527	88262512 (3053, 3054)	Novel Protein sim. GBank gi 2792496 (AF041107) - lulip 2 (Rattus norvegicus)			56182575, 264259, 60432049, 29331822, 60432289, 264908, 66712502, 60433438, 87168559, 265017, 264288, 21906766, 21908789, 263977, 55811576, 56182323, 18108381
1528	94130818 (3055, 3056)			UNCLASSIFIED	22278995, 22278997, 264259, 66712502, 264596, 265017, 265019, 264682, 264448, 264683, 264764, 264685, 264686, 21908765, 21906766, 21906787, 21906788, 21908769, 265022, 264693, 83373044, 18108385
1529	94120793 (3057, 3058)	Novel Protein sim. GBank gi 4406663 gb AAD20053  - (AF131826) Unknown (Homo sapiens)		UNCLASSIFIED	264488, 263994, 56182575, 22278995, 35695286, 22278997, 264259, 29331822, 60432289, 29331827, 35696052, 264509, 264906, 264907, 264908, 264909, 52844045, 56182435, 264511, 265009, 264910, 60433356, 60433438, 265017, 265018, 264780, 264448, 264764, 264369, 264288, 264768, 18108357, 264788, 52844229, 21905765, 21908766, 21908787, 21908788, 265021, 265022, 52844150, 33657109, 264629, 35695855, 60432113, 22279002, 264583, 264564, 264485, 264567
1530	95012765 (3059, 3060)	Novel Protein sim. GBank gi 2828710 (AF043642) - matrin cytophtin (Rattus norvegicus)			264488, 264489, 35696286, 29331825, 35696052, 264508, 264905, 264906, 264907, 264908, 264510, 264511, 264512, 264910, 264592, 264593, 18108351, 264764, 264683, 264684, 264766, 264768, 18108357, 264769, 35695917, 264628, 264629, 18108374, 35695855, 264630, 264631, 264634, 264555, 264636, 264637, 264404, 264563, 264568, 264486
1531	95419351 (3061, 3062)	Novel Protein sim. GBank gi 1905874 (U90878) - carboxyl terminal LIM domain protein (Homo sapiens)	Contains protein domain (PF00595) - kinase PDZ domain (Also known as DHR or GLGF).		56182575, 35696286, 264097, 264259, 29331822, 29331825, 29331826, 29331827, 35696052, 264508, 56182435, 264510, 264511, 265007, 60433356, 55811386, 264681, 264369, 264288, 264766, 264687, 55811957, 35695917, 33657023, 35695763, 55810764, 35696423, 55811576, 263981, 60170394, 56182323, 83373044, 60432113, 264568

1532	85718224 (3063, 3064)	Novel Protein sim. GBank gi 3874716 emb CAA91265  - (Z68484) cDNA EST EMBL:D65271 comes from this gene; cDNA EST EMBL:D64845 comes from this gene; cDNA EST EMBL:D64448 comes from this gene; cDNA EST EMBL:D67438 comes from this gene; cDNA EST EMBL:D68087 comes from this gene; cDN...			UNCLASSIFIED	264689
1533	94239830 (3065, 3066)	Novel Protein sim. GBank gi 1490324 emb CAB01543  - (Z78141) unknown [Mus musculus]			structural	29331824, 29148489, 264907, 264112, 265008, 265011, 265017, 265018, 264762, 18108351, 263967, 20281149, 18108374, 263981, 264566
1534	95343941 (3067, 3068)	Novel Protein sim. GBank gi 51286 pir J522897 - extensin - Valvex carteri (fragment)			UNCLASSIFIED	264905, 264907, 264766, 264637
1535	90936732 (3069, 3070)					65274572, 22278987, 264259, 60432049, 29331822, 60432288, 29331827, 28148489, 265006, 265008, 60170831, 60433438, 33109954, 87168559, 265018, 18108357, 21908766, 29148629, 265021, 265022, 18108377, 58182323, 60432113, 22279000, 22279002
1536	87602856 (3071, 3072)	Novel Protein sim. GBank gi 106024 pir B32891 - finger protein 2, placental - human	Contains protein domain (PF00095) - Zinc finger, C2H2 type	transcription factor		264686, 18108357, 18108394, 21906767, 21806768, 29148629, 35696286, 265020, 265021, 52644150, 264693, 68714117, 265021, 2644150, 264693, 68714117, 29331825, 29331826, 264508, 264905, 20281148, 284909, 18108374, 35696423, 35695855, 265009, 264634, 264636, 264638, 18108385, 56528486, 285017, 285018, 264563, 264762, 18108351, 264448, 264369, 264766
1537	96354556 (3073, 3074)	Novel Protein sim. GBank gi 3876332 emb CAB02096  - (Z79754) cDNA EST EMBL:T01054 comes from this gene; cDNA EST EMBL:D73600 comes from this gene; cDNA EST yk426112.5 comes from this gene; cDNA EST yk342110.5 comes from this gene; cDNA ES... comes from this gene; cDNA ES...				65274572, 56182575, 60432049, 264259, 29331826, 265006, 265007, 60433356, 60433438, 264601, 18108351, 264448, 264369, 264288, 33657023, 65274620, 33657109, 60432113
1538	85724628 (3075, 3076)	Novel Protein sim. GBank gi 403440 (M81787) - [Gallus domesticus skeletal muscle mRNA, partial cds.] gene product [Gallus gallus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase		18108394, 18108387, 264609, 265008, 265009, 265010, 18108351, 284938, 18108382, 18108385, 18108388



1539	95337828 (3077, 3078)	Novel Protein sim. GBank gij321841 [emb/CAA19575.1] - (AL023859) SPBC19C7.07c, putative tRNA splicing endonuclease gamma subunit, len:284aa, similar eg. to YAR008W, YAH8_YEAST, P39707, YAR008W, tRNA splicing endonuclease gamma subunit, (275aa), fasta scores, opt:269, E(1):6.e-2...	nuclease	22278994, 22278996, 35696286, 56994075, 22278997, 22278998, 22278999, 60432049, 264258, 29331822, 29331824, 60424269, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 33656970, 56182435, 265008, 33657402, 60433356, 60433438, 55812038, 21906754, 87168558, 285017, 265018, 265019, 18108351, 264288, 52844228, 18108359, 21906764, 21906787, 21906768, 35695917, 265020, 265021, 52844150, 33657023, 33657109, 27465261, 18108370, 18108376, 35696423, 55811576, 65274791, 284558, 56182323, 60170394, 83373044, 87168518, 60432113, 22279000, 22279002, 264588
1540	95352858 (3079, 3080)	Novel Protein sim. GBank gij5052834 [gb/AA038647.1] AF14567 - (AF145672) BcDNA GH12174 [Drosophila melanogaster]	UNCLASSIFIED	264369, 264691, 263978
1541	95317948 (3081, 3082)	Novel Protein sim. GBank gij5052349 [gb/AA038515.1] AF13501 - (AF135016) protein phosphatase 2A 48 kDa regulatory subunit [Homo sapiens]	phosphatase	264488, 264489, 22278998, 264259, 29331822, 35696052, 264508, 264509, 264905, 264907, 264908, 284511, 284512, 264910, 264592, 264761, 264762, 264448, 264764, 264288, 264687, 21906769, 55811957, 35695917, 265020, 264691, 33657023, 264692, 33657109, 264628, 18108374, 264632, 264634, 264635, 264639, 18108385, 264583, 264584, 264585, 264586, 264485
1542	90937549 (3083, 3084)	Novel Protein sim. GBank gij5305702 [gb/AA041779.1] AF12686 - (AF126867) calpain-like protease [Mus musculus]	calthepsin	18108392, 18108394, 65274572, 29331822, 264508, 265007, 265008, 265009, 265011, 264692, 18108394, 18108395, 52644150, 18108398, 264636, 18108381, 18108382
1543	84348768 (3085, 3086)	Novel Protein sim. GBank gij728832 [sp/P39189] ALU2_HUMAN - III ALU SUBFAMILY SB WARNING ENTRY IIII	nuclease	264108, 33108934, 265019, 264683, 35695917, 284690, 284692, 33657109
1544	87757285 (3087, 3088)	Novel Protein sim. GBank gij3493162 (AF084259) - bromodomain-containing protein BP75 [Mus musculus]	Contains protein domain (PF00439) - nucl_rept Bromodomain	35696286, 56994075, 22278998, 35696052, 60433356, 60433438, 265011, 264683, 33657109, 35696423, 264631, 87168518, 22279000
1545	85757973 (3089, 3090)	Novel Protein sim. GBank gij1086591 (U41007) - similar to S. cerevisiae nuclear protein SNF2 (SP-P22082) in a region of 61y-arg repeats [Caenorhabditis elegans]	UNCLASSIFIED	264112, 264692, 264693, 55811576
1546	79476589 (3091, 3092)		UNCLASSIFIED	264905, 264688
1547	86999584 (3093, 3094)	Novel Protein sim. GBank gij2661132 (AF035683) - p21 [Mus musculus]	UNCLASSIFIED	264259, 29331822, 66714117, 265007, 55811386, 265010, 264600, 265017, 265018, 264288, 264768, 265020, 265022, 55811576, 18108380, 264563

1548	94233065 (3095, 3096)	Novel Protein sim. GBank gij3043692[dbj BAA25510] - (AB011159) KIAA0584 protein [Homo sapiens]		UNCLASSIFIED	29331824, 60431528, 264539, 56182323
1549	95330048 (3097, 3098)	Novel Protein sim. GBank gij5689519[dbj BAA83043.1] - (AB029014) KIAA1091 protein [Homo sapiens]		eph	60424179, 22278995, 35656286, 22278998, 22278999, 264092, 264094, 29331822, 56182181, 29331824, 35688052, 284905, 284906, 284908, 284909, 285008, 284511, 265008, 60431735, 60433356, 21908754, 55811386, 87168559, 285017, 285018, 265019, 55811150, 284682, 284288, 264369, 56181562, 264769, 21906765, 21906768, 21906769, 55811957, 265020, 264691, 33657109, 60431528, 35696423, 35695855, 56526486, 60432113, 22279002, 264563, 264566
1550	95201907 (3099, 3100)	Novel Protein sim. GBank gij544463[sp P35350 GUSB_BOVIN - POSSIBLE GUSTATORY RECEPTOR TYPE B (PPR1 PROTEIN)]	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	Im7	65274572, 60432289, 265008, 264910, 285011, 265017, 265019, 284768, 56182323
1551	88077111 (3101, 3102)	Novel Protein sim. GBank gij4758566[ref NP_004798.1 pH56S - heparan-sulfate 6-sulfotransferase		UNCLASSIFIED	22278999, 29331822, 264508, 264509, 264905, 264907, 284909, 265007, 284512, 264910, 21906754, 265018, 265019, 284681, 284764, 264766, 264688, 264769, 21906769, 284692, 35693763, 284635, 264535, 264556, 284537, 264638, 264558, 264563
1552	87617114 (3103, 3104)			UNCLASSIFIED	284288, 284688, 33657109, 264556, 56182575, 35688286, 28146499, 264509, 264907, 284908, 284909, 56182435, 265006, 265008, 285009, 264910, 284757, 264758, 265017, 55811150, 18108351, 284764, 56181562, 35695917, 284693, 33657109, 18108374, 35696423, 65274791, 35695855, 264635, 264555, 56182323, 18108392, 83373044, 22279000
1553	94725512 (3105, 3106)	Novel Protein sim. GBank gij4588370[dbj BAA76807.1] - (AB023180) KIAA0963 protein [Homo sapiens]	Contains protein domain (PF00304) - Gamma-thionins family	- dehydrogenase	35698286, 22278997, 284259, 29331822, 29331824, 29331825, 29331828, 265007, 265009, 60432229, 33657402, 55812038, 265011, 265019, 264681, 264369, 264686, 264767, 264768, 21906765, 21906769, 35695917, 284693, 18108370, 60431528, 55811576, 264631, 60170394, 56182323, 83373044, 18108385, 22279000, 22279002
1554	94233069 (3107, 3108)	Novel Protein sim. GBank gij3043692[dbj BAA25510] - (AB011159) KIAA0584 protein [Homo sapiens]	Contains protein domain (PF00446) - Gonadotropin-releasing hormones	transferase	

1555	87332970 (3109, 3110)	Novel Protein sim. GBank gi 2257495 dbj BAA21392  - (AB004534) p015 [Schizosaccharomyces pombe]	Contains protein domain (PF00400) - UNCLASSIFIED WD domain, G-beta repeat	264259, 29331826, 35696052, 264508, 264905, 264908, 264907, 264908, 52844045, 264910, 60432229, 60433356, 55812038, 264758, 264759, 33637084, 265011, 87168559, 264601, 265018, 265019, 264763, 264764, 264288, 264766, 264768, 21906765, 35695917, 265022, 264691, 33657023, 35696423, 35695855, 264635, 264555, 264638, 264638, 264639, 18108385, 56526486
1556	91228268 (3111, 3112)		UNCLASSIFIED	83373044, 264758, 265022, 264600, 35696052, 264630, 35696423, 265018, 264632, 264682, 29331822, 265020, 265011, 60432289, 264508, 264908, 264907, 264908, 264909, 264910, 264758, 55811386, 264761, 264762, 264766, 264769, 264690, 263978, 264634, 264635, 264639, 264584, 264488
1557	67840609 (3113, 3114)	Novel Protein sim. GBank gi 3329511 (AF078783) - contains similarity to C3HC4-type zinc fingers (Pfam; zf-C3HC4.hmm, score: 34.06); most similar to dirosophila galliath protein (SW: Q06003) [Caenorhabditis elegans]	Contains protein domain (PF00087) - UNCLASSIFIED Zinc finger, C3HC4 type (RING finger)	22278994, 22278996, 22278997, 22278998, 22278999, 60432048, 264259, 29331824, 29331825, 29331826, 29331827, 264908, 264909, 60433356, 21906767, 265021, 265022, 265018, 264448, 21906767, 265021, 265022, 33657023, 33657109, 18108370, 55811578, 83373044, 87168558, 22278900, 22279002
1558	94840376 (3115, 3116)	Novel Protein sim. GBank gi 5360105 gb AA042871.1 AF15510 - (AF155105) putative zinc finger protein NY-REN-34 antigen [Homo sapiens]		264259, 29331822, 60432289, 35696032, 264107, 264110, 21906754, 33108954, 87168559, 264760, 264763, 21906764, 21906765, 21906769, 265021, 264690, 35695855
1559	88224865 (3117, 3118)	Novel Protein sim. GBank gi 112808 sp P02750 AZGL_HUMAN - LEUCINE-RICH ALPHA-2-GLYCOPROTEIN (LRG)	Contains protein domain (PF00560) - glycoprotein Leucine Rich Repeat	264259, 29331822, 60432289, 35696032, 264107, 264110, 21906754, 33108954, 87168559, 264760, 264763, 21906764, 21906765, 21906769, 265021, 264690, 35695855
1560	84580675 (3119, 3120)	Novel Protein sim. GBank gi 3880146 emb CAA92704  - (Z68319) Similarity to Human hnRNP F protein (PIR Acc. No. S43484); cDNA EST EMBL:D34219 comes from this gene; cDNA EST EMBL:D37248 comes from this gene; cDNA EST EMBL:D71817 comes from this gene; cDNA EST EMBL:D74531 comes from...	UNCLASSIFIED	264908, 264603, 264638
1561	86609159 (3121, 3122)		UNCLASSIFIED	264510, 264594
1562	83359682 (3123, 3124)		UNCLASSIFIED	263987
1563	85508694 (3125, 3126)		UNCLASSIFIED	264910, 264764, 264766
1564	87766371 (3127, 3128)	Novel Protein sim. GBank gi 1168287 sp P45953 ACDV_RAT - ACYL-COA DEHYDROGENASE, VERY-LONG-CHAIN SPECIFIC PRECURSOR (VLCD)	Contains protein domain (PF00441) - dehydrogenase Acyl-CoA dehydrogenase	18108394, 35698286, 264259, 29331822, 60432288, 35696052, 29331828, 264508, 58712502, 264908, 56182435, 265007, 264910, 60170831, 21906754, 265011, 265017, 265018, 265019, 18108351, 264448, 264288, 264688, 21906765, 21906768, 265021, 60170815, 264692, 35696423, 35695855, 264557, 56182323, 60432113, 22279002, 264482

1565	87783381 (3129, 3130)	Novel Protein sim. GBank gij129726[sp]P05307[PDI_BOVIN] - PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI) (PROLYL 4-HYDROXYLASE BETA SUBUNIT) (CELLULAR THYROID HORMONE BINDING PROTEIN) (P55)		isomerase	264488, 264689, 18103398, 55811857, 264534, 264259, 264508, 264905, 264509, 264908, 18108372, 264510, 264511, 264512, 265008, 264630, 265009, 264910, 264635, 264636, 264591, 264555, 264592, 264637, 264593, 264594, 264595, 264596, 265011, 264603, 22279002, 18108351, 264782, 264555, 264567
1566	87424748 (3131, 3132)	Novel Protein sim. GBank gij388044[emb]CAA20329] - (AL031266) VM106R.1 [Caenorhabditis elegans]		int	22278996, 22278998, 264259, 29331822, 29331824, 60432289, 29331827, 66712502, 264806, 265008, 18108351, 52644229, 21906765, 21906767, 21906768, 21906769, 33657109, 264555, 264639, 264482
1567	84999008 (3133, 3134)	Novel Protein sim. GBank gij492989[gb]AAD34110.1[AF15187] - (AF151873) CGI-115 protein [Homo sapiens]		UNCLASSIFIED	56182575, 21908769, 264692
1568	87848761 (3135, 3136)	Novel Protein sim. GBank gij4827063[re]NP_005072.1[pZNF1 - zinc finger protein 142 (clone pHZ-49)]	Contains protein domain (PF000096) - Zinc finger, C2H2 type	transcript factor	29331827, 29331830, 264511, 265009, 264788, 21906767, 21906768, 264691, 264693, 22279000, 22279002
1569	90836668 (3137, 3138)	Novel Protein sim. GBank gij5689451[dbj]BAA83009.1] - (AB028980) KIAA1057 protein [Homo sapiens]	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	65274572, 29331822, 29331824, 29331828, 264905, 56182435, 265007, 265018, 264784, 21906765, 21906769, 55811957, 60170615, 52644150, 264692, 33657023, 33657109, 18108377, 264583, 264567
1570	86943981 (3139, 3140)	Novel Protein sim. GBank gij1255430 [U53155] - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	264595, 264692, 265021
1571	91210340 (3141, 3142)	Novel Protein sim. GBank gij4507731[re]NP_001081.1[pTUBG - tubulin, gamma polypeptide]	Contains protein domain (PF000091) - Tubulin/FtsZ family	tubulin	22278998, 35698286, 22278997, 264091, 264259, 29331824, 29331825, 29331827, 35698032, 264508, 264905, 56182435, 264510, 265007, 264758, 265011, 18108351, 264448, 264288, 264369, 21906765, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 33657023, 264693, 18108370, 18108377, 35698423, 35695855, 264634, 264555, 264558, 18108384
1572	81201664 (3143, 3144)			UNCLASSIFIED	35698032, 264905, 264906, 264908, 264910, 264758, 264768, 35695917, 264637
1573	80207066 (3145, 3146)			UNCLASSIFIED	263972

1574	94216142 (3147, 3148)	Novel Protein sim. GBank gij4756334/reiNP_004256.1 pFADS - delta-6 fatty acid desaturase	Contains protein domain (PF00173) - Heme-binding domain in cytochrome b5 and oxidoreductases	cytochrome	18108394, 264867, 18108397, 18108398, 22278990, 22278997, 22278999, 264239, 29331825, 29331827, 29148498, 29148499, 264107, 264907, 264909, 52644045, 264511, 265008, 264910, 265009, 264591, 21908754, 265011, 265018, 18108351, 264682, 264763, 264784, 18108354, 264368, 264288, 264685, 264766, 264688, 264768, 264688, 21908785, 21908786, 21908787, 21908788, 21908789, 29148829, 264690, 264691, 264693, 20281069, 18108370, 18108374, 18108378, 35695855, 264634, 18108384, 18108385, 22279002, 264583, 264586
1575	95340018 (3149, 3150)	Novel Protein sim. GBank gij3881810 emb CAA94856 - (Z70783) similar to EF-hand calcium binding protein; cDNA EST EMBL C08700 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00036) - EF hand	phosphatase	56594075, 264259, 29331822, 29331824, 29331825, 60432289, 29331828, 264909, 265006, 265008, 265009, 265010, 87168559, 55811150, 264448, 18108354, 264369, 264288, 18108357, 55811957, 265020, 265021, 60170615, 264691, 33657023, 33657109, 60431528, 65274791, 35695855, 18108385, 60432113, 22279002, 264482
1576	95314019 (3151, 3152)	Novel Protein sim. GBank gij2773195 (AF039711) - contains similarity to Physcomitrella patens glyceraldehyde 3-phosphate dehydrogenase (GB:X72381) [Caenorhabditis elegans]		UNCLASSIFIED	264588, 264082, 264094, 264095, 264259, 264508, 264905, 264509, 264907, 264909, 264510, 264511, 265008, 264910, 21908754, 265010, 265011, 87168559, 264761, 264762, 264288, 264766, 264769, 264691, 264693, 35695855, 264632, 264634, 264635, 264636, 83373044, 264486
1577	87613800 (3153, 3154)	Novel Protein sim. GBank gij2495130 sp P70315 WASP_MOUSE - WISKOTT- ALDRICH SYNDROME PROTEIN HOMOLOG [WASP]		UNCLASSIFIED	264511, 265011, 264681, 264369, 264686, 264689, 264629, 264555, 264558, 264559
1578	87123138 (3155, 3156)				264259, 29331826, 265017, 264689, 264693, 60432113
1579	88085141 (3157, 3158)	Novel Protein sim. GBank gij2978255 dbj BAA251901 - (AB007407) myeloid zinc finger protein-2 [Mus musculus]	Contains protein domain (PF00086) - Zinc finger, C2H2 type	UNCLASSIFIED	35696286, 264908, 264909, 60433438, 55811386, 264388, 264685, 33657023, 264555, 264556, 264557, 87168518
1580	87255702 (3159, 3160)	Novel Protein sim. GBank gij4324682 gb AAD169861 - (AF108874) lata gestation lung protein 1 [Rattus norvegicus]	Contains protein domain (PF00188) - SCP-like extracellular protein	glycoprotein	22278998, 35696052, 29331830, 52644045, 55812038, 87168474, 265018, 264448, 265022, 264638, 56526485, 22279000
1581	95087431 (3161, 3162)	Novel Protein sim. GBank gij2088838 (AF003385) - F59E12.4 gene product [Caenorhabditis elegans]			22278995, 29331822, 29331824, 29331826, 56182435, 264595, 55812038, 87168559, 265017, 264288, 21908764, 55811957, 35695917, 264692, 55811576, 264637, 56182323, 264559, 83373044, 60432113

1582	95358052 (3163, 3164)	Novel Protein sim. GBank gij5420387[embjCAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]		phosphatase	264259, 60432289, 29331827, 264509, 264905, 264906, 264907, 264909, 264910, 264782, 264288, 264768, 264769, 264632, 264555, 264639, 56525486, 22275000, 60170831, 33657402, 284682, 21908768, 35695855, 284563
1583	87622715 (3165, 3166)	Novel Protein sim. GBank gij5578958[embjCAB51351.1] - (AL050306) dJ475B7.2 (novel protein) [Homo sapiens]		UNCLASSIFIED	
1584	95337722 (3167, 3168)	Novel Protein sim. GBank gij5531815[gbjAAD44482.1] - (AF078850) steroid dehydrogenase homolog [Homo sapiens]	Contains protein domain (PF00106) - short chain dehydrogenase	dehydrogenase	60424179, 52646842, 65274572, 56182575, 22278995, 35686286, 22278996, 22278998, 22278999, 264259, 29331822, 56182181, 60424269, 60432289, 29331827, 29331828, 35696052, 29146498, 66712502, 29331830, 52644045, 56182435, 284510, 284512, 285008, 60433356, 33657402, 60433438, 55812038, 21906754, 55811386, 52644296, 87188474, 87168559, 265018, 265019, 264448, 264389, 264288, 18108358, 21906765, 21906787, 21906788, 21906769, 35695917, 265020, 265021, 285022, 52644150, 33657023, 33657109, 18108374, 55810764, 55811578, 35686423, 65274781, 35695855, 56182323, 83373044, 18108387, 87188518, 50432113, 22279002
1585	87626117 (3169, 3170)	Novel Protein sim. GBank gij4240132[dbjBAA74846.1] - (AB020630) KIAA0823 protein [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	phosphatase	35686286, 22278998, 264259, 29331822, 29331824, 29331825, 264905, 285008, 265007, 265008, 60433356, 33108854, 87168474, 285011, 265017, 264604, 264369, 264288, 264685, 264789, 18108359, 21906765, 18108364, 18108370, 264629, 263972, 18108383, 18108388, 264482, 264564
1586	88067081 (3171, 3172)	Novel Protein sim. GBank gij3786494 (AF098993) - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	265017, 265018, 264689, 33657023, 263978, 264636, 264563
1587	87617126 (3173, 3174)	Novel Protein sim. GBank gij3253159 (AF005355) - translation initiation factor eIF2C [Oryzotagus cuniculus]		UNCLASSIFIED	264907, 264908, 264511, 284910, 264591, 284594, 284628, 284631, 264563, 284483, 284587
1588	87802536 (3175, 3176)	Novel Protein sim. GBank gij1077573[priJIS52680 - probable ribosomal protein L34, mitochondrial - yeast] (Saccharomyces cerevisiae)	Contains protein domain (PF00468) - Ribosomal protein L34	UNCLASSIFIED	264259, 29331828, 264905, 265008, 264758, 21906754, 264761, 264762, 21906785, 21906769, 60170815, 52644150, 33637108, 35695855, 56182323, 18108385
1589	80880653 (3177, 3178)	Novel Protein sim. GBank gij2137756[priJ148746 - semaphorin C - mouse (fragment)]		UNCLASSIFIED	65274572, 264480, 29331822, 66714117, 29331827, 29331828, 56182435, 265008, 60170831, 264595, 264758, 264596, 285011, 264686, 21906786, 21906768, 55811957, 27486265, 264639, 18108385, 56526486, 60432113

1590	95319825 (3179, 3180)			UNCLASSIFIED	264489, 22278986, 264259, 29331824, 29331825, 29331826, 29331827, 265006, 60433356, 21906754, 285017, 265018, 265019, 284448, 264765, 264288, 52844229, 21906765, 21906767, 21906768, 21906769, 265021, 264692, 27480285, 35893783, 56526488, 60432113, 22279000, 22279002, 264564
1591	86877160 (3181, 3182)	Novel Protein sim. GBank		MHC	264259, 264905, 29331830, 264595, 265017, 284448, 284288, 284690, 264628, 87168518
1592	87882533 (3183, 3184)	gi 4557749 ref NP_000237.1 pMHC2 - MHC class II transactivator			
1593	94891561 (3185, 3186)			UNCLASSIFIED	65274572, 60432049, 284509, 60433356, 21906754, 21906767, 21906768, 18108370, 35696423, 22279000, 264565, 264567
1594	87773752 (3187, 3188)	Novel Protein sim. GBank gi 3877072 emb CAA87060  (Z46937) similarity with ribosomal protein L21 [Caenorhabditis elegans]		UNCLASSIFIED	264488, 29331827, 264905, 264906, 264907, 284908, 284909, 264910, 264592, 264593, 264757, 284602, 284604, 264760, 264681, 264288, 264766, 284768, 29148829, 35695917, 284692, 264628, 264629, 264630, 264632, 264634, 264635, 264636, 264639, 284583, 264584, 264586
1595	78018425 (3189, 3190)	Novel Protein sim. GBank gi 3152703 (AF065389) - tetraspan NET-4 [Homo sapiens]	Contains protein domain (PF00335) - 4 transmembrane segments integral membrane proteins	UNCLASSIFIED	29331826, 284908, 55811957
1596	79933928 (3191, 3192)	Novel Protein sim. GBank		UNCLASSIFIED	29146498, 264758, 263987
1597	86971957 (3193, 3194)	gi 5257114 gb AAD41244.1 AF094480 cholesterol 24-hydroxylase [Homo sapiens]	Contains protein domain (PF00067) - Cytochrome P450	cyto450	264092, 29331824, 284508, 284682, 284389, 264686, 264630, 264563
1598	87862939 (3195, 3196)	Novel Protein sim. GBank			264259, 264634
1599	87649829 (3197, 3198)	gi 4505797 ref NP_000324.1 pSCA7 - spinocerebellar ataxia 7 (olivopontocerebellar atrophy with retinal degeneration)	Contains protein domain (PF01581) - FMRFamide related peptide family	UNCLASSIFIED	52845080, 29331824, 29331826, 264511, 285009, 265011, 284605, 264448, 264764, 265020, 284692, 284693, 18108370, 264635, 18108385
1600	80056002 (3199, 3200)	Novel Protein sim. GBank		UNCLASSIFIED	29331826, 264603, 284681, 284583
1601	15023248 (3201, 3202)				284635
1602	86926987 (3203, 3204)	Novel Protein sim. GBank gi 5305704 gb AAD41780.1 AF12853 - (AF128535) cytoplasmic phosphoprotein PACSIN2 [Mus musculus] chicken	Contains protein domain (PF00018) - SH3 domain	sinud	29146499, 264112, 264762, 18108351, 29148627, 263974
1603	80502072 (3205, 3206)	Novel Protein sim. GBank gi 283920 pir I S27839 - lensin - chicken		collagen	264480, 28331824, 264907, 264909, 264511, 265008, 284592, 285010, 285011, 264762, 264764, 284369, 284288, 284687, 284769, 264693, 264628, 264634, 264636, 264555, 284556, 264638, 264557, 264558, 264559, 18108385
1604	80221813 (3207, 3208)	Novel Protein sim. GBank gi 4768831 gb AAD29633.1 AF11682 - (AF116827) unknown [Homo sapiens]		ATPase-associated	263977

1605	91221129 (3209, 3210)				264905, 264509, 264906, 264907, 264908, 264809, 264804, 264766, 264768, 264692, 264693, 33657109, 264629, 35695855, 264635, 264636, 264637	struct	
1606	94312703 (3211, 3212)	Novel Protein sim. GBank gi 4505313 ref NP_003764.1 p MYOM - UNKNOWN	Contains protein domain (PF00047) - Immunoglobulin domain		22278996, 22278999, 264259, 33657402, 265017, 18108351, 264448, 21908767, 21906769, 52644150, 264691, 87168518	struct	
1607	10871805 (3213, 3214)	Novel Protein sim. GBank gi 5174473 ref NP_005888.1 p PPI - intracisternal A particle-promoted polypeptide			264689	transcriptfactor	
1608	80428900 (3215, 3216)	Novel Protein sim. GBank gi 2224629 dbj BAA20802 - (AB002342) KIAA0344 [Homo sapiens]			264094, 264906, 264907, 264909, 264910, 264591, 264603, 264768, 264693, 264634, 264635, 264637, 264639	UNCLASSIFIED	
1609	94311572 (3217, 3218)	Novel Protein sim. GBank gi 4884073 emb CA843213.1 - (AL049934) hypothetical protein [Homo sapiens]			52644507, 52645156, 52646365, 52648842, 56182575, 22278994, 56684075, 35696288, 22278997, 22278998, 22278999, 264259, 52645080, 28147820, 29331826, 35696052, 33656970, 264508, 264509, 264907, 52644045, 56182435, 264510, 264511, 264512, 33657402, 21906754, 52646317, 33109954, 52644296, 87168474, 265017, 265018, 265019, 18108351, 264448, 264288, 264769, 52644228, 21906765, 21906768, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 52644150, 33657023, 33657109, 52645129, 27486261, 27486262, 35695763, 264628, 18108370, 18108376, 35696423, 264638, 52644332, 18108387, 87168516, 22279000, 264563, 264486		
1610	55468200 (3219, 3220)	Novel Protein sim. GBank gi 263920 pir S27939 - lensin - chicken			264593, 264757, 55812038, 265018, 265020, 264691, 264692, 264693, 264631, 264634, 264635, 264555, 22279000, 264564	UNCLASSIFIED	
1611	94122843 (3221, 3222)	Novel Protein sim. GBank gi 107284 pir A35415 - peroxidase (EC 1.11.1.7), thyroid (Grave's disease) - human (fragment)	Contains protein domain (PF00008) - EGF-like domain		35696286, 21906765, 264691, 35696423	peroxidase	
1612	85746031 (3223, 3224)	Novel Protein sim. GBank gi 3874846 emb CAA94337 - (Z70307) Similarity to B. subtilis tetracycline resistance protein (SW.TCR2_BACSU); cDNA EST EMBL:CO9951 comes from this gene; cDNA EST EMBL:CO8265 comes from this gene [Caenorhabditis elegans]			264488, 264508, 18108370, 18108387, 264486	UNCLASSIFIED	
1613	82247354 (3225, 3226)				264759	UNCLASSIFIED	



1614	91228634 (3227, 3228)	Novel Protein sim. GBank gi 4680673 gb AA027726.1 AF132951 - (AF132951) CGI-17 protein [Homo sapiens]	Contains protein domain (PF01605) - ERF-1-like proteins	UNCLASSIFIED	22278995, 22278996, 22278997, 22278998, 22278999, 284258, 29331822, 284908, 284512, 285009, 265011, 265017, 265018, 265019, 18108351, 264683, 264288, 264766, 21906787, 21906788, 21906789, 35895917, 265021, 265022, 35896423, 35895855, 60170394, 56182323, 83373044, 264566, 22278995, 35896286, 22278997, 29331822, 35896052, 29331828, 284508, 284908, 284909, 56182435, 264511, 265017, 265018, 264766, 264767, 284768, 265020, 264691, 284628, 264632, 264635, 264555, 264556, 56182323, 264558, 22279002
1615	86121909 (3229, 3230)	Novel Protein sim. GBank gi 5689485 gb BAA63026.1  - (AB028987) KIAA1074 protein [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	homeobox	264488, 52644507, 52645156, 52646365, 52646842, 22278994, 22278995, 35898286, 22278996, 22278997, 22278999, 52645080, 29331822, 29331824, 29331825, 29331827, 29331828, 35896052, 35896870, 264905, 264909, 264594, 52646317, 21906754, 33857084, 52644286, 87168474, 87168558, 265017, 265018, 285019, 284681, 284448, 284684, 52644229, 21906784, 284689, 21908785, 21908786, 21908789, 35895917, 265020, 265021, 52644150, 33857023, 52645129, 33857109, 33857182, 27486261, 27486262, 33857349, 27486265, 35895783, 18108376, 35896423, 35895855, 264557, 52644332, 264558, 18108385, 87168518
1616	94311819 (3231, 3232)	Novel Protein sim. GBank gi 3876260 emb CA801696  - (Z78418) cDNA EST EMBL:D71020 comes from this gene; cDNA EST EMBL:D73593 comes from this gene; cDNA EST EMBL:C07849 comes from this gene; cDNA EST EMBL:C09081 comes from this gene; cDNA EST yk399/2.3 comes from this gene; cDNA ...	UNCLASSIFIED	UNCLASSIFIED	35896052, 264805, 284509, 264907, 264908, 284510, 264511, 284784, 284786, 284788, 284689, 264693, 18108374, 264635, 264636, 264638
1617	88030742 (3233, 3234)	Novel Protein sim. GBank gi 466053 sp P34679 YO41_CAEEL - HYPOTHETICAL 68.7 KD PROTEIN ZK757.1 IN CHROMOSOME III	Contains protein domain (PF01529) - DHHC zinc finger domain	peptidase	35896286, 22278999, 264092, 29331824, 29331825, 35896052, 33857084, 21906765, 27486264
1618	86272860 (3235, 3236)	Novel Protein sim. GBank gi 424023 gb BAA74894.1  - (AB020678) KIAA0871 protein [Homo sapiens]		strud	52646842, 65274572, 22278999, 284259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 35896052, 56182435, 285007, 265008, 264910, 60170831, 60432229, 60433356, 60433438, 265019, 284448, 284288, 284686, 21906768, 265021, 60170815, 33857023, 65274620, 33857109, 18108374, 18108376, 35896423, 35895855, 56182323, 56526488
1619	95354580 (3237, 3238)	Novel Protein sim. GBank gi 5031763 ref NP_005515.1 pHRY - hairy (Drosophila)- homolog	Contains protein domain (PF00010) - Helix-loop-helix DNA-binding domain	transcriptfactor	264684
1620	87344655 (3239, 3240)	Novel Protein sim. GBank gi 1351047 sp P45843 SCRT_DROME - SCARLET PROTEIN		UNCLASSIFIED	

1621	87076708 (3241, 3242)	Novel Protein sim. GBank gi 3822553 (AF098788) - nuclear calmodulin-binding protein [Gallus gallus]	Contains protein domain (PF00622) - SPRY domain	UNCLASSIFIED	284910	18108392, 65274573, 18108398, 22278996, 22278997, 22278999, 29146498, 29146499, 284905, 284908, 284909, 284928, 52644045, 284592, 60433356, 21806754, 284602, 285017, 284359, 21908788, 55811957, 285021, 60170615, 284635, 284557, 60170394, 83373044, 18108385, 22279000, 22279002, 284566
1622	84741739 (3243, 3244)	Novel Protein sim. GBank gi 731086 sp P40389 UV22_SCHPO - UV-INDUCED PROTEIN UV122		ribosomalprot		18108398, 284259, 284909, 56182435, 87168474, 264448, 21905768, 35695917, 284691, 87168518, 264583
1624	87338178 (3247, 3248)	Novel Protein sim. GBank gi 3875665 emb CAB05478  (Z83104) cDNA EST EMBL: T00015 comes from this gene; cDNA EST EMBL: D33653 comes from this gene; cDNA EST EMBL: D36340 comes from this gene; cDNA EST yk2408.3 comes from this gene; cDNA EST yk38768.3 comes from this gene; cDNA ES...		UNCLASSIFIED		68714117, 29331825, 284908, 265008, 264758
1625	95354748 (3249, 3250)	Novel Protein sim. GBank gi 4589622 db BAA76833.1  (AB023206) KIAA0989 protein [Homo sapiens]		kinase		264489, 22278994, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 264092, 264259, 29331824, 29331825, 29331827, 29331828, 264102, 264108, 284508, 33657084, 265017, 265018, 18108351, 264683, 264369, 264288, 21908785, 21908786, 21908787, 21908789, 35695917, 285021, 264691, 65274620, 18108368, 263972, 18108376, 35696423, 264631, 284634, 22279000, 22279002
1626	94734399 (3251, 3252)	Novel Protein sim. GBank gi 567907 gb AA048844.1 AF16080 - (AF160804) BcDNA HL05936 [Drosophila melanogaster]				52644507, 52645156, 52646385, 52648842, 22278994, 56994075, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 66714117, 29331826, 60432289, 29331827, 29331828, 35696052, 264906, 66712502, 264809, 265008, 265009, 60432228, 60433356, 60433438, 21905754, 52646317, 52644298, 285011, 87168559, 284604, 265018, 284448, 284369, 264286, 264766, 52644229, 264889, 21806785, 21908788, 35695917, 285021, 285022, 52644150, 33657023, 65274620, 27486261, 27486282, 27486285, 35695783, 263972, 52644332, 60170384, 87168518, 60432113, 264567
1627	83368773 (3253, 3254)	Novel Protein sim. GBank gi 3568087 (AC004667) - hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED		264288
1628	85708459 (3255, 3256)			eph		264288, 264586, 264767, 22279002

1629	84693841 (3257, 3258)	Novel Protein sim. GBank gij4240175dbj BAAT4866.1  - (AB020650) KIAA0843 protein [Homo sapiens]		struct	264555
1630	87778027 (3259, 3260)			UNCLASSIFIED	29331822, 29331827, 265010, 264693, 264634, 22279002
1631	87758454 (3261, 3262)	Novel Protein sim. GBank gij1915892 emb CAA69995  - (Y08740) lom-1A protein [Gallus gallus]		UNCLASSIFIED	55811957, 264259, 33657023, 264693, 29331822, 29331824, 29331827, 29331828, 264908, 264908, 55811576, 264910, 264634, 264636, 264637, 58182323, 264539, 264758, 18108385, 264563, 264764, 264768
1632	87871692 (3263, 3264)	Novel Protein sim. GBank gij2558501 dbj BAA22886  - (D83850) hepatoma-derived growth factor [Mus musculus]		UNCLASSIFIED	264687, 264769, 264691, 264692, 29146489, 264509, 264905, 264907, 264511, 264512, 264482, 264681, 264763, 264682, 264683, 264488, 264259, 264907, 264908, 264909, 264628, 264629, 264631
1633	87773683 (3265, 3266)				265007, 264637, 22279002
1634	85992817 (3267, 3268)	Novel Protein sim. GBank gij4887229 gb AAD32244.1 AF150755 - (AF150755) microtubule-actin crosslinking factor [Mus musculus]	Contains protein domain (PF00435) - Spectrin repeat	struct	
1635	94232600 (3269, 3270)			UNCLASSIFIED	65274572, 22278998, 35988052, 52844045, 264511, 265008, 265009, 265010, 265011, 265018, 265019, 264448, 264369, 21908765, 21908768, 265021, 264690, 264482
1636	80413227 (3271, 3272)	Novel Protein sim. GBank gij455751 ref NP_001339.1 pDAPK - death-associated protein kinase 3	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	UNCLASSIFIED	22278995, 264594, 264763, 265020
1637	80070435 (3273, 3274)			kinase	264558
1638	87101854 (3275, 3276)	Novel Protein sim. GBank gij3420051 (AC004680) - unknown protein [Arabidopsis thaliana]			21908765, 21908767, 22278996, 35986286, 22278998, 264259, 264692, 264693, 29331824, 33657109, 264508, 264906, 18108370, 264628, 265007, 33657402, 21908754, 264602, 264604, 264764, 264683, 264566, 264288
1639	94322194 (3277, 3278)	Novel Protein sim. GBank gij5420389 emb CAB46680.1  - (AJ243460) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264488, 18108394, 65274572, 56182575, 35986286, 29331824, 29331826, 29331827, 35986082, 264908, 56182435, 264112, 265006, 265008, 264757, 264758, 55811386, 264603, 264760, 18108351, 264764, 264288, 264786, 264788, 21908767, 55811957, 264691, 33657023, 65274620, 18108370, 55810764, 55811576, 264558, 264639, 83373044, 18108385, 87168518

1640	84143185 (3279, 3280)	Novel Protein sim. GBank gi 2842459 emb CAA16847.1 - (AL021747) hypothetical protein [Schizosaccharomyces pombe]		UNCLASSIFIED	56182575, 56894075, 35686286, 60432049, 60432288, 29331827, 35696052, 52644045, 56182435, 264510, 265006, 265007, 265008, 264910, 265009, 33657402, 55812038, 265010, 265011, 265017, 265018, 264288, 52644229, 21908785, 21908788, 21908789, 35695817, 265021, 60170815, 52644150, 33657023, 33657109, 33657349, 18108374, 35696423, 65274791, 35695855, 264632, 264555, 56182323, 22279000
1641	87625160 (3281, 3282)			UNCLASSIFIED	29146499, 265006, 265007, 265008, 265009, 55812038, 265010, 265011, 264555, 264558, 264558, 18108383
1642	94312557 (3283, 3284)	Novel Protein sim. GBank gi 1575333 (U60416) - myr 6 myosin heavy chain [Rattus norvegicus]	Contains protein domain (PF01843) - struct DIL domain		22278999, 29147620, 28331826, 28331828, 33656970, 55812038, 265010, 265018, 265019, 18108351, 264889, 265020, 265022, 264690, 33657023, 65274620, 35695763, 52644332, 18108381, 60170394, 56182323, 18108388, 87168518, 22279002, 264584
1643	94131766 (3285, 3286)				29331825, 29331827, 29331828, 21908754, 265019, 264288, 264893, 33657349, 18108370, 18108378, 264555, 83373044, 22278002, 264482
1644	88095125 (3287, 3288)			UNCLASSIFIED	264905, 264907, 264908, 264910, 265009, 264757, 264758, 264761, 264782, 264783, 264766, 264768, 264769, 264828, 264829, 264630, 264631, 264632, 264553, 264564, 264585, 264588, 264587
1645	95013658 (3289, 3290)			UNCLASSIFIED	264685, 264693
1646	95362691 (3291, 3292)	Novel Protein sim. GBank gi 1076802 pi S49915 - extensin like protein - maize		UNCLASSIFIED	22278994, 56994075, 35696266, 264259, 29331824, 29331825, 29331826, 60432289, 264508, 60433356, 60433438, 87168559, 265018, 264687, 35695917, 264692, 33657023, 33657182, 27486261, 27486265, 33657349, 60432113, 264583, 264584
1647	94278428 (3293, 3294)	Novel Protein sim. GBank gi 5002573 emb CAB44338.1 - (Y17466) alpha-N-acetylglucosamine alpha-2,6- sialyltransferase [Fugu rubripes]		UNCLASSIFIED	29331822, 264906, 264908, 264369, 21908788, 60170815, 264639, 22279000
1648	87642098 (3295, 3296)		Contains protein domain (PF00096) - Zinc finger, C2H2 type	UNCLASSIFIED	265009, 264686, 55811957, 35695917, 55810764, 264556, 56182323, 264558, 18108385

1649	95347628 (3297, 3298)	Novel Protein sim. GBank gij854065[emb]CAA583371 - (X83413) U88 [Human herpesvirus 6]		cadherin	264488, 22278995, 35896286, 22278996, 22278997, 22278998, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264905, 264907, 66712502, 264908, 52644045, 264909, 56182435, 264511, 265007, 265008, 265009, 264591, 264593, 60433438, 264598, 55812038, 21906754, 265011, 264501, 264602, 265017, 265018, 265019, 264682, 264448, 264764, 264683, 264288, 264766, 264685, 264687, 264768, 264688, 264769, 52644229, 264689, 21906765, 21908788, 21906767, 21906768, 55811957, 35695917, 265021, 265022, 52644150, 264892, 33657023, 33657109, 20281148, 18108370, 264628, 18108374, 18108376, 35696423, 35695855, 264632, 264634, 264635, 264636, 18108380, 264639, 264558, 18108382, 18108384, 18108385, 18108387, 264080, 264404, 60432113, 22279000, 22279002, 264482, 264565, 264566, 264487, 265011, 264802, 21906767, 18108374, 18108377, 18106385
1650	67418539 (3299, 3300)	Novel Protein sim. GBank gij3647335[emb]CAA210591 - (AL031644) possible zinc-finger protein [Schizosaccharomyces pombe]			
1651	91639773 (3301, 3302)	Novel Protein sim. GBank gij4894278[emb]CAB43247.1 - (AL050037) hypothetical protein [Homo sapiens]		synthase	264488, 52645156, 18108397, 35696286, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264508, 264908, 29331830, 264910, 60432228, 21906754, 265010, 265011, 265017, 265019, 264448, 18108354, 264286, 264688, 21906765, 21906766, 21906768, 21906769, 265022, 264692, 264693, 264629, 35695855, 264556, 264637, 264557, 264559, 63373044, 56526486, 22279000, 22279002, 264584
1652	86598622 (3303, 3304)	Novel Protein sim. GBank gij1657837 (U73200) - p16Rip [Mus musculus]	Contains protein domain (PF00169) - PH domain	siucl	22278997, 26146488, 56182435, 21906754, 264389, 21906765, 21906766, 21906769, 265020, 52644150, 33657109, 22279000, 22279002
1653	94255893 (3305, 3306)	Novel Protein sim. GBank gij3776054[emb]CAA062731 - (AJ004989) Tapasin [Gallus gallus]	Contains protein domain (PF00047) - immunoglobulin domain	glycoprotein	18108398, 22278995, 22278998, 264259, 29331822, 29331824, 60432289, 29331828, 29331827, 29331830, 264809, 265006, 265009, 60432228, 60433356, 60433438, 21906754, 265017, 265018, 264448, 264683, 264288, 265021, 265022, 264692, 18108364, 65274791, 18108384, 60432113, 264567, 33657109, 264565
1654	79756471 (3307, 3308)			UNCLASSIFIED	

1655	86689346 (3308, 3310)	Novel Protein sim. GBank gij3355717[emb CAA73496] - (Y13053) seryl-tRNA synthetase [Zea mays]		synthase	52644507, 35696286, 22278998, 22278999, 29331824, 29331825, 29331826, 33656970, 284908, 52644045, 264511, 264910, 52646317, 284288, 52644229, 33657023, 33657109, 52644332, 264557, 56182323, 56528486, 60432113
1656	78962287 (3311, 3312)	Novel Protein sim. GBank gij1890141[db BAA18947] - (D83206) P24 protein [Mus musculus]		UNCLASSIFIED	29331822, 29331824, 29331825, 264563
1657	87771994 (3313, 3314)	Novel Protein sim. GBank gij4537645[ref NP_001524.1]pHNRP - heterogeneous nuclear ribonucleoprotein L		Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	29331827, 265009, 21906766, 21906767, 265020, 265022, 33657109, 264638, 56528486, 264482
1658	87773778 (3315, 3316)	Novel Protein sim. GBank gij3877072[emb CAA87060] - (Z46937) similarity with ribosomal protein L21 [Caenorhabditis elegans]		UNCLASSIFIED	52646355, 35696286, 22278998, 22278999, 29331824, 29331825, 29331826, 29331827, 29331828, 29148498, 284905, 284908, 52644045, 265005, 60433336, 264757, 60433438, 21906764, 265011, 18108351, 264448, 264389, 264288, 264766, 264768, 21906785, 21906787, 21906768, 21906769, 29148829, 265021, 265022, 18108362, 263989, 263971, 18108374, 35698423, 18108383, 22279000, 284482
1859	88230101 (3317, 3318)	Novel Protein sim. GBank gij539218[pir S38038] - hypothetical protein YKL201c - yeast (Saccharomyces cerevisiae)		UNCLASSIFIED	52646317, 21906766, 21906767, 21906768, 87168518, 22278998, 265020, 22278999, 87168559, 264603, 265017, 264631, 265018, 265019, 22279002, 264482, 264635, 264565
1860	94315313 (3318, 3320)	Novel Protein sim. GBank gij2497012[sp Q10010]YSV4, CAEL - HYPOTHETICAL 26.6 KD PROTEIN T19C3.4 IN CHROMOSOME III		UNCLASSIFIED	264488, 35696286, 284259, 35696032, 264508, 284509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265008, 265007, 264512, 265009, 264910, 264592, 264596, 265010, 264600, 264602, 265017, 265018, 264605, 264760, 264764, 264288, 284766, 284686, 284768, 284769, 264689, 21906786, 35698423, 284690, 33657023, 284693, 33657108, 264638, 35698423, 35698555, 264634, 264635, 264555, 264638, 264637, 264556, 264638, 264639, 264559, 18108385, 18108386, 264563, 264483, 264584, 264585, 264586, 264486, 264567

1661	94234071 (3321, 3322)	Novel Protein sim. GBank gi 4759100 ref NP_004759.1 pSFRS - splicing factor, arginine/serine-rich 11	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264488, 2227898, 264258, 26331824, 29331828, 29331827, 29331828, 264509, 66712502, 29331830, 264908, 52844045, 265007, 264512, 60433356, 60433438, 55812038, 21906754, 265019, 264448, 284768, 264768, 264769, 21906768, 21906768, 265020, 33657023, 33657109, 65274791, 87168518, 264482, 264563, 264564, 264565, 264567
1662	94135172 (3323, 3324)	Novel Protein sim. GBank gi 1730502 sp P52875 PE27_MOUSE - TRANSMEMBRANE PROTEIN PF127			18108392, 29331822, 29331828, 20281100, 284108, 265008, 265007, 265008, 18108348, 21906768, 18108385, 18108386, 18108374, 83373044, 18108385
1663	94217146 (3325, 3326)	Novel Protein sim. GBank gi 4884138 emb CAB43275.1  - (AL050107) hypothetical protein [Homo sapiens]	Contains protein domain (PF00397) - WW domain	kinase	52845156, 56182575, 22278994, 22278995, 35696288, 22278996, 59994075, 22278997, 22278998, 22278999, 264259, 29331822, 29331828, 29331827, 29331828, 33655970, 29331830, 264908, 56182435, 264511, 60433356, 33657402, 33109954, 87168474, 87168559, 265017, 265018, 264605, 18108351, 284764, 284288, 264766, 264768, 21906755, 21906768, 21906767, 21906768, 21906769, 265021, 265022, 264691, 33657023, 264693, 263987, 33657109, 264630, 52644332, 83373044, 87168518, 60432113, 22278000
1664	94234076 (3327, 3328)	Novel Protein sim. GBank gi 3043682 dbj BAA25510  - (AB011156) KIAA0584 protein [Homo sapiens]		UNCLASSIFIED	264488, 263984, 35696288, 29331824, 35696052, 264508, 264509, 284905, 284908, 264907, 264908, 264909, 264510, 264511, 265009, 264910, 60170831, 264591, 264592, 264595, 87168474, 265011, 264600, 264601, 264604, 264605, 264760, 264762, 18108351, 264681, 264682, 264763, 264683, 264764, 284288, 264684, 264766, 264687, 264768, 264769, 21906764, 21906765, 21906767, 35695917, 265021, 264534, 60170615, 284690, 264691, 264692, 33657109, 33657182, 264628, 18108370, 264629, 35696423, 35695855, 264634, 264635, 284555, 264636, 264637, 264638, 264639, 264558, 83373044, 87168518, 264563, 264566, 264486

1665	91226952 (3329, 3330)	Novel Protein sim. GBank gll1083506[pi]IS0065 - sialoadhesin - mouse	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	264488, 29331826, 29331828, 264509, 264906, 264907, 264909, 264510, 264511, 264910, 264592, 264593, 264595, 264758, 264596, 264600, 264760, 264762, 264764, 264766, 264768, 264629, 264630, 264634, 264636, 83373044, 264564, 264566, 264567, 264486
1666	95358160 (3331, 3332)	Novel Protein sim. GBank gll3913431[sp]O42643[DDX8_SCHPO - PUTATIVE PRE- MRNA SPLICING FACTOR ATP-DEPENDENT RNA HELICASE SPAC10F6.02C	Contains protein domain (PF00575) - S1 RNA binding domain	helicase	56994075, 22278999, 264359, 29331824, 29331826, 29331827, 29146498, 265009, 33109954, 87168559, 265019, 264288, 264688, 21906767, 21906769, 264691, 33657182, 18108370, 18108374, 18108385, 22278002
1667	91228655 (3333, 3334)	Novel Protein sim. GBank gll5689535[dl]BAA83051.1] - (AB029022) KIAA1089 protein [Homo sapiens]	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	transport	264259, 29331822, 29331826, 264905, 264906, 264908, 264510, 285009, 264595, 264758, 265011, 87168559, 265017, 265018, 265019, 264448, 264766, 264686, 21906765, 21906767, 21906769, 265020, 265021, 60170615, 264690, 264692, 264693, 18108368, 18108370, 263972, 55810764, 264555, 83373044, 60432113, 22279000, 22279002
1668	88095135 (3335, 3336)	Novel Protein sim. GBank gll2076894[gl]BAA83983.1] - (AF002197) short region of weak similarity to protein kinase C; contains similarity to Pfam domain PF00130 (DAG PE- bind). Score=10.0, E-value=0.0034, N=1 [Caenorhabditis elegans]		kinase	66714117, 264508, 264509, 264906, 264907, 264908, 264511, 264810, 264784, 264887, 264689, 33657109, 35696423, 35695855, 264632
1669	91227846 (3337, 3338)	Novel Protein sim. GBank gll3875371[emb]CAA85414.1] - (Z36948) contains a valine and arginine rich domain, possesses weak similarity with the RNA binding domains from RNA splicing factor U2AF 65 KD subunit; cDNA EST EMBL:D64658 comes from this gene; cDNA EST EMBL:D66829 comes fr...		UNCLASSIFIED	29331825, 33109954, 264369, 264767, 264689, 33657109, 83373044
1670	87828009 (3339, 3340)			UNCLASSIFIED	264259, 29331824, 29331827, 60433438, 265022, 264838
1671	87346372 (3341, 3342)	Novel Protein sim. GBank gll482451[sp]P3424[KKK1_YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL101W	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	35696286, 22278997, 29331825, 264909, 21906754, 265017, 265018, 265019, 264682, 264683, 264768, 264688, 21906766, 21906767, 21906768, 21906769, 264691, 264555, 264558, 22278000, 264568
1672	86291834 (3343, 3344)	Novel Protein sim. GBank gll1814270 (U74586) - double- stranded RNA specific adenosine deaminase [Rattus norvegicus]		deaminase	264906, 264909, 264632, 18108381



1673	86095137 (3345, 3348)	Novel Protein sim. GBank gjl2076894[gbjAAB53983.1] - (AF002197) short region of weak similarity to protein kinase C; contains similarity to Pfam domain PF00130 (DAG_PIE-bind). Score=10.0, E-value=0.0034, N=1 [Caenorhabditis elegans]	Contains protein domain (PF00130) - Phorbol esters/diacylglycerol binding domain (C1 domain)	kinase	264488, 264569, 18108394, 56994075, 22278998, 264256, 35696052, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264510, 264910, 60170831, 264592, 264594, 264595, 264758, 264601, 264760, 264762, 264683, 264784, 264288, 264788, 264688, 264788, 264687, 264789, 264689, 264690, 33657023, 264692, 264693, 33657109, 264628, 264629, 18108374, 35696423, 35695855, 264631, 264632, 264634, 264635, 264637, 264556, 264638, 264636, 264583, 264482, 264584, 264565, 264566, 264567, 264488
1674	88258028 (3347, 3348)	Novel Protein sim. GBank gjl5262467[embjCAB45693.1] - (AL080062) hypothetical protein [Homo sapiens]		kinase	29331822, 29331824, 264908, 5264045, 80433356, 87168559, 264448, 264288, 264688, 264691
1675	87606466 (3349, 3350)	Novel Protein sim. GBank gjl3128368 (AF010496) - 50S ribosomal protein l9 [Rhodobacter capsulatus]		UNCLASSIFIED	56181686, 35696286, 22278997, 22278998, 264259, 29331824, 29331827, 35696052, 66712502, 264764, 264288, 264686, 264687, 35695917, 265020, 264690, 264693, 35695763, 18108370, 35696423, 35695855, 264637, 264639, 18108385, 264584
1676	95358086 (3351, 3352)	Novel Protein sim. GBank gjl4184065[gbjAAD053271] - (AF111091) latrophilin 3 splice variant bba1 [Bos taurus]		UNCLASSIFIED	264259, 29331827, 29331828, 264106, 264907, 265009, 264600, 265019, 264288, 21906765, 265020, 265022, 35695855, 83373044, 18108385
1677	87408587 (3353, 3354)	Novel Protein sim. GBank gjl3327046[dbjBAA31591] - (AB014516) KIAA0616 protein [Homo sapiens]		UNCLASSIFIED	264908
1678	86866829 (3355, 3356)	Novel Protein sim. GBank gjl550452 (U08469) - 3-methylcrotonyl-CoA carboxylase, biotin-carrier domain [Glycine max]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPSase)	UNCLASSIFIED	29331824, 264102
1679	91214108 (3357, 3358)	Novel Protein sim. GBank gjl550452 (U08469) - 3-methylcrotonyl-CoA carboxylase, biotin-carrier domain [Glycine max]		carboxylase	264488, 18108392, 18108394, 52646842, 18108397, 18108398, 35696286, 29331824, 265006, 265007, 265008, 265009, 18108348, 265011, 18108351, 264683, 18108354, 18108358, 18108359, 21906765, 29148627, 29148629, 264690, 18108361, 18108362, 18108364, 18108365, 18108368, 264628, 18108379, 35696423, 35695855, 264635, 18108381, 18108382, 18108383, 18108384, 18108385, 18108388
1680	81005372 (3359, 3360)	Novel Protein sim. GBank gjl2394478 (AF024500) - No definition line found [Caenorhabditis elegans]		transport	85274572, 22278994, 22278999, 66714117, 29331827, 56182435, 21906754, 265018, 264288, 21906769
1681	84324150 (3361, 3362)	Novel Protein sim. GBank gjl5689537[dbjBAA83052.1] - (AB029023) KIAA1100 protein [Homo sapiens]		UNCLASSIFIED	22278996, 29331822, 264908, 264593, 264604, 265019, 264683, 55811957, 264690, 33657023, 35696423, 83373044, 264563
1682	86042710 (3363, 3364)			UNCLASSIFIED	264909, 265017, 264505

1683	94316213 (3365, 3366)	Novel Protein sim. GBank gi 5031717 ref NP_005704.1 pGPBP - goodpasture antigen- binding protein	Contains protein domain (PF01832) - UNCLASSIFIED START domain	263994, 35696286, 35696032, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 265006, 265007, 265009, 264910, 264593, 264758, 265010, 265011, 264760, 264761, 264762, 264763, 264764, 264368, 264766, 264768, 35695917, 264692, 33657109, 35698423, 35695855, 264634, 264635, 264636, 264638, 264639, 83373044, 264486
1684	80063409 (3367, 3368)	Novel Protein sim. GBank gi 1255371 (U53147) - coded for by C. elegans cDNA yk34a9.5; coded for by C. elegans cDNA yk34a9.3; Similar to guanylate kinase. [Caenorhabditis elegans]	UNCLASSIFIED Contains protein domain (PF00628) Guanylate kinase	264563, 264566 60424179, 52646842, 22278994, 35696286, 22278998, 264259, 52645080, 29331824, 29331826, 265007, 33657084, 265018, 264681, 264448, 264683, 264369, 264689, 21908765, 21908787, 21908789, 21908789, 265021, 264692, 65274620, 33657109, 27486262, 264635, 52644332, 56182323, 22279000
1686	87620710 (3371, 3372)	Novel Protein sim. GBank gi 2244707 dbj BAA21115.1  - (AB005287) thrombospondin 1 [Bos taurus]	UNCLASSIFIED	56182575, 264258, 264508, 264905, 264509, 264907, 264908, 264510, 264511, 265006, 264512, 265009, 264910, 264758, 265010, 265011, 264805, 18108351, 264764, 264766, 18108357, 264768, 18108362, 264628, 264630, 264631, 264634, 264635, 264637, 264638, 264639, 264565, 264486, 264567
1687	94719400 (3373, 3374)	Novel Protein sim. GBank gi 4680879 gb AA027729.1 AF13295 - (AF132954) CGI-20 protein [Homo sapiens]	UNCLASSIFIED	33695917, 264906, 264907, 264908, 264510, 265006, 265007, 264910, 264558, 18108381, 18108383, 285011
1688	82158442 (3375, 3376)	Novel Protein sim. GBank gi 4240193 dbj BAA74875.1  - (AB020659) KIAA0852 protein [Homo sapiens]	UNCLASSIFIED	264905, 264910, 264760, 264629, 264555
1689	94325049 (3377, 3378)	Novel Protein sim. GBank gi 4240193 dbj BAA74875.1  - (AB020659) KIAA0852 protein [Homo sapiens]	UNCLASSIFIED	264559, 35696286, 22278998, 264508, 264908, 264909, 56182435, 33657402, 60433438, 55812038, 285017, 265018, 264448, 264764, 264288, 264686, 29148629, 35695917, 265020, 265021, 263972, 18108374, 65274791, 83373044, 264089
1690	83255346 (3379, 3380)	Novel Protein sim. GBank gi 3800736 (AF031572) - seven- pass transmembrane receptor precursor [Mus musculus]	Contains protein domain (PF00028) - cadherin Cadherin domain	264369, 21908766, 264692, 264639, 87188518
1691	88095223 (3381, 3382)	Novel Protein sim. GBank gi 2773208 (AF039713) - No definition line found [Caenorhabditis elegans]		264768, 33657109, 29331827, 29148629, 264510, 264106, 264910, 264109, 264508, 60170831, 264563, 264905, 264564, 264691, 264637, 264628, 264907, 264908, 33657023, 264587, 264768, 263974
1692	86106709 (3383, 3384)	Novel Protein sim. GBank gi 121271 sp P02207 GLB_LAMEL - GLOBIN	UNCLASSIFIED	264106
1693	87012775 (3385, 3386)	Novel Protein sim. GBank gi 121271 sp P02207 GLB_LAMEL - GLOBIN	Contains protein domain (PF00042) - UNCLASSIFIED Globin	29331826, 264508, 264905, 264907, 264595, 265010, 265011, 21908768, 33657023, 264629, 263978, 264558

1694	94208168 (3387, 3388)	Novel Protein sim. GBank gi15453932[refINP_006225.1]pPOLR - polymerase (RNA) II (DNA directed) polypeptide J (13.3kD)	Contains protein domain (PF01193) - RNA polymerases L / 13 to 16 kDa subunit	mapolymerase	35696286, 22278998, 22278998, 22278999, 264259, 29331822, 29331825, 29331826, 29331828, 35696052, 28148499, 264805, 284908, 52644045, 264511, 265006, 265007, 265008, 264592, 60433356, 21908754, 265010, 265011, 18108351, 264763, 264682, 284448, 264883, 264288, 264788, 264689, 21906768, 60170815, 264691, 264692, 264693, 18108370, 18108374, 263976, 35696423, 35695855, 264556, 18108381, 18108385, 87168518, 264482, 264486, 264634
1695	84719325 (3389, 3390)	Novel Protein sim. GBank gi14680879[gbAAD27729.1]AF132954 CGI-20 protein [Homo sapiens]	UNCLASSIFIED		
1696	87824038 (3391, 3392)	Novel Protein sim. GBank gi14220517[emb]CAA229301 - (AL035356) hypothetical protein [Arabidopsis thaliana]			22278997, 264259, 265010, 18108351, 264784, 21908766, 18108370
1697	85740563 (3393, 3394)	Novel Protein sim. GBank gi1505652 (U10382) - GP36b glycoprotein [Homo sapiens]	glycoprotein		264682
1698	87445285 (3395, 3396)	Novel Protein sim. GBank gi15052031[gbAAD38411.1]AF15573 - (AF155739) axolotrin [Mus musculus]			56994075, 22278998, 22278999, 264509, 33557402, 264758, 87168474, 87168559, 285017, 265018, 264448, 264687, 29148627, 21908769, 28148629, 265020, 265022, 33657023, 264558, 87168518, 22279002
1699	87424783 (3397, 3398)	Novel Protein sim. GBank gi15052031[gbAAD38411.1]AF15573 - (AF155739) axolotrin [Mus musculus]	UNCLASSIFIED		35696286, 264635
1700	87659161 (3399, 3400)	Novel Protein sim. GBank gi1543344[pir]S1647 - zinc finger 5 protein - mouse	Contains protein domain (PF00098) - Zinc finger, C2H2 type	nucl_rept	29331824, 52644045, 265008, 265009, 263969, 263971
1701	86570488 (3401, 3402)	Novel Protein sim. GBank gi13877439[emb]CAA966521 - (Z72510) similarity to yeast UTR3 protein (Swiss Prot accession number P21374); cDNA EST EMBL:D72822 comes from this gene; cDNA EST EMBL:D75763 comes from this gene; cDNA EST YK274e3.3 comes from this gene; cDNA EST YK274e3.5 c...	UNCLASSIFIED	MHC	264092, 264110, 263977
1702	87755092 (3403, 3404)	Novel Protein sim. GBank gi13877439[emb]CAA966521 - (Z72510) similarity to yeast UTR3 protein (Swiss Prot accession number P21374); cDNA EST EMBL:D72822 comes from this gene; cDNA EST EMBL:D75763 comes from this gene; cDNA EST YK274e3.3 comes from this gene; cDNA EST YK274e3.5 c...			22278995, 22278997, 264092, 29146488, 29146499, 264107, 264508, 264807, 264110, 264112, 265009, 60170831, 21906754, 265011, 265017, 264762, 18108351, 264288, 21906765, 35695917, 265021, 60170615, 263967, 33657109, 18108370, 263972, 263974, 18108374, 263976, 35695855, 264555, 263981, 60170394, 18108385, 56526486, 87168518, 60432113
1703	79568651 (3405, 3406)	Novel Protein sim. GBank gi1451544 (U04267) - proline-rich cell wall protein [Gossypium barbadense]	UNCLASSIFIED		264909, 265017, 264628, 264629, 264638
1704	86622979 (3407, 3408)	Novel Protein sim. GBank gi1263289 (U47856) - fibrin-4 [Araneus diadematus]	UNCLASSIFIED		264369
1705	87795175 (3409, 3410)	Novel Protein sim. GBank gi14519621[db]BAA75670.11 - (AB017814) OASIS protein [Mus musculus]			264559, 35696286, 264907, 265010, 264687, 264768, 264692, 264693, 264636, 264568
1706	87750967 (3411, 3412)	Novel Protein sim. GBank gi13123034[sp]Q15011Y025_HUMAN - HYPOTHETICAL PROTEIN KIAA0025			22278998, 22278998, 264259, 264509, 265018, 264784, 264685, 264686, 21906768, 21906769, 265022, 264691, 264558, 22279000

1707	88041230 (3413, 3414)	Novel Protein sim. GBank gij4321664[gb AAD15797] - (AF055470) ZNF258 [Homo sapiens]		UNCLASSIFIED	18108396, 22278997, 264259, 29147620, 29331826, 29146498, 264905, 264908, 285008, 264493, 264595, 264758, 264596, 265018, 264780, 18108351, 264764, 264766, 264689, 264693, 18108370, 35696423, 55811576, 264558, 87168518, 60432113, 264567
1708	91220519 (3415, 3416)	Novel Protein sim. GBank gij5174591[ref NP_005947.1 pMTfH - 5,10-methylenetetrahydrofolate dehydrogenase, 5,10-methylenetetrahydrofolate cyclodiolase, 10-formyltetrahydrofolate synthetase	Contains protein domain (PF01288) - Formate-tetrahydrofolate ligase	synthase	56182575, 22278996, 56994075, 264259, 29331822, 29331824, 29331826, 29331827, 29331828, 29146498, 29146499, 29331830, 265009, 60170831, 33657402, 33109854, 87168559, 265019, 18108351, 264448, 21808765, 21808787, 21808788, 29148627, 29148628, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264558, 18108385, 22279000, 264563
1709	80222583 (3417, 3418)			UNCLASSIFIED	264107, 55811957, 263974, 263976, 263977, 263981
1710	20754572 (3419, 3420)			UNCLASSIFIED	264556
1711	91013729 (3421, 3422)	Novel Protein sim. GBank gij5031735[ref NP_005760.1 pHEC - N-acetylglucosamine 6-O-sulfotransferase		sulfotransferase	65274572, 29331824, 29331826, 264768, 60431528, 35696423, 60432113, 264563
1712	95330184 (3423, 3424)	Novel Protein sim. GBank gij5454168[ref NP_008453.1 pXAP4 - HBV associated factor	Contains protein domain (PF00641) - Kinase Zn-finger in Ran binding protein and others.	kinase	56994075, 264093, 264259, 29331822, 264099, 29331824, 29331827, 264107, 264110, 264511, 264592, 265011, 265018, 264683, 264686, 264689, 265020, 33657023, 263967, 33657109, 263974, 35696423, 35695855, 264630, 264636, 264558, 264566
1713	84143453 (3425, 3426)	Novel Protein sim. GBank gij160409 (M69183) - mature-parasite-infected erythrocyte surface antigen [Plasmodium falciparum]	Contains protein domain (PF00643) - B-box zinc finger.	UNCLASSIFIED	22278995, 264508, 264758, 18108351, 18108370, 263974, 18108374, 264634, 56182323, 83373044, 60432113
1714	87420048 (3427, 3428)				22278997, 264757, 21806785, 265020, 265021, 264692, 56526486
1715	94260257 (3429, 3430)	Novel Protein sim. GBank gij5689537[dbj BA83052.1  - (AB029023) KIAA1100 protein [Homo sapiens]		UNCLASSIFIED	264509, 264905, 264906, 264907, 264908, 264909, 264910, 264591, 265011, 264766, 264768, 264769, 264691, 264692, 264632, 264634, 264635, 264636, 264637, 264558, 264639, 264564
1716	87400448 (3431, 3432)	Novel Protein sim. GBank gij4589468[dbj BA76761.1  - (AB012808) mBOCT [Mus musculus]		transport	56182575, 29331824, 60432288, 264109, 264809, 285007, 264600, 265019, 264886, 265020, 264693, 55811578, 264558, 60432113, 22279002

1717	87563223 (3433, 3434)	Novel Protein sim. GBank gij2765411[emb]CAA74749 - [Y14391] GTP-binding protein [Homo sapiens]		UNCLASSIFIED	264569, 264259, 29331825, 29331826, 29331828, 35996052, 264509, 264905, 264907, 264908, 264909, 264512, 265009, 264910, 264592, 264595, 264758, 264759, 265017, 264681, 264764, 264766, 264686, 18108357, 35895917, 264690, 264692, 264693, 264628, 264629, 35696423, 264630, 264631, 264635, 264636, 18108380, 264638, 264639, 18108389, 18108391
1718	87032628 (3435, 3436)	Novel Protein sim. GBank gij2633262[sp]Q14999[Y076] HUMAN - HYPOTHETICAL PROTEIN KAA0076 (HA0936)		UNCLASSIFIED	265011, 264681, 264682, 264684, 264688, 264689, 21906765, 265021, 264691, 33657023, 284693, 18108370, 35695855, 264632, 264634, 264636, 18108388, 22279002
1719	94315259 (3437, 3438)	Novel Protein sim. GBank gij4505187[ref]NP_003473.1[pMLL2 - myeloid/lymphoid or mixed-lineage leukemia 2		UNCLASSIFIED	16108396, 65274572, 35698286, 22278987, 60432049, 59182181, 66714117, 60432289, 29331826, 35996052, 29331828, 264906, 29331830, 56182435, 264592, 60431735, 60433438, 55812038, 264759, 265010, 264600, 264601, 265017, 264448, 264764, 264288, 264768, 21906766, 21906769, 33657023, 33657109, 33657182, 27486282, 55811957, 265020, 265021, 52844150, 18108374, 35898423, 55811576, 35695855, 33657349, 35695783, 18108370, 60431528, 264631, 56182323, 264559, 264584, 264486, 56182575, 22278989, 264259, 29331824, 60432289, 29331827, 35686052, 264509, 264905, 264906, 264907, 264608, 264909, 264511, 264910, 264758, 21906754, 265011, 264601, 264760, 264762, 264288, 264766, 264686, 18108357, 264689, 21906765, 55811957, 264693, 20281149, 264629, 18108374, 55811576, 65274791, 264630, 20281071, 264634, 264635, 264636, 264637, 264558, 264638, 264639, 56182323, 87168518
1720	84653063 (3439, 3440)	Novel Protein sim. GBank gij2129478[p]j551939 - chitinase (EC 3.2.1.14) precursor - bee		UNCLASSIFIED	264631, 56182323, 264559, 264584, 264486, 56182575, 22278989, 264259, 29331824, 60432289, 29331827, 35686052, 264509, 264905, 264906, 264907, 264608, 264909, 264511, 264910, 264758, 21906754, 265011, 264601, 264760, 264762, 264288, 264766, 264686, 18108357, 264689, 21906765, 55811957, 264693, 20281149, 264629, 18108374, 55811576, 65274791, 264630, 20281071, 264634, 264635, 264636, 264637, 264558, 264638, 264639, 56182323, 87168518
1721	91722288 (3441, 3442)	Novel Protein sim. GBank gij4886451[emb]CAB43381.1 - (AL050280) hypothetical protein [Homo sapiens]		UNCLASSIFIED	22278994, 22278999, 29331822, 265006, 265007, 265008, 55812038, 21906754, 60174639, 265011, 87168559, 18108351, 18108354, 21906785, 21906766, 21906768, 21906769, 265020, 33657109, 18108370, 18108374, 264556, 60170394, 83373044, 18108385, 264486
1722	94134549 (3443, 3444)	Novel Protein sim. GBank gij5689375[db]BAA82968.1 - (AB030644) tudor repeat associator with PCTAIRE 2 [Rattus norvegicus]	Contains protein domain (PF00567) - kinase Tudor domain		55594075, 29331824, 29331828, 265009, 18108351, 21906768, 265020, 33657023, 18108374, 83373044

1723	95358181 (3445, 3446)	Novel Protein sim. GBank gi 4426962 gb AAD20533  - (AF126082) Arf-like 2 binding protein BART1 [Homo sapiens]	UNCLASSIFIED	264486, 264687, 264769, 21906767, 21906768, 56182575, 55811957, 22278997, 22278998, 265020, 264259, 264692, 33857023, 29331822, 29331824, 29331825, 60432289, 33857182, 33856970, 33857349, 29146499, 264508, 264907, 18108370, 264629, 264808, 264909, 18108374, 55811576, 264510, 265008, 264511, 265007, 264910, 264632, 264591, 60432228, 264592, 60433356, 264594, 60433438, 264595, 83373044, 55812038, 33108954, 33857084, 87168518, 87168474, 265010, 265011, 87168558, 264600, 80432113, 264604, 285019, 264593, 264448, 264682, 264586, 264764, 264288, 264567, 264488, 264389, 264766
1724	87713806 (3447, 3448)	Novel Protein sim. GBank gi 2340182 (AF005083) - dsRBP-ZF a [Xenopus laevis]	UNCLASSIFIED	264905, 18108359, 264693, 264628, 264631, 264636, 264555, 264556, 264558, 264559
1725	85655191 (3449, 3450)	Novel Protein sim. GBank gi 3152562 (AF064604) - KE03 protein [Homo sapiens]	homeobox	35686286, 264259, 29331822, 35686052, 264508, 264509, 264905, 264906, 264907, 264908, 264809, 264910, 265009, 264591, 264601, 264760, 18108351, 264681, 264764, 264288, 264766, 264788, 21906768, 35685917, 264628, 35686423, 264630, 264631, 264632, 264635, 264638, 264638, 87168518, 264566
1726	85754255 (3451, 3452)	Novel Protein sim. GBank gi 4689348 gb AAD27861.1 AF13256 - (AF132562) BcDNA.LD14270 [Drosophila melanogaster]	UNCLASSIFIED	29146498, 264683, 264689
1727	85286382 (3453, 3454)		UNCLASSIFIED	264905, 265011, 264689, 21906768

1728	85346515 (3455, 3456)	Novel Protein sim. GBank gij4406549jpb AA020027  - (AF131738) Unknown [Homo sapiens]		UNCLASSIFIED	60424179, 18108397, 56182575, 22278995, 56994075, 35696286, 22278997, 22278998, 22278999, 264094, 60432049, 264259, 29331822, 29331824, 56182181, 29331825, 60432289, 29331826, 29331827, 35696052, 264905, 264906, 264907, 29331830, 66712502, 264808, 59182435, 264511, 285008, 265008, 60432228, 60433356, 33857402, 60433438, 264759, 21908754, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 55811150, 264681, 264448, 264682, 264763, 264683, 264288, 264684, 264389, 264685, 264786, 264687, 264769, 21908784, 264689, 21906765, 21908766, 21906767, 21906768, 35695917, 265020, 265021, 265022, 264535, 264691, 264692, 33657023, 264693, 33657109, 18108370, 264628, 263972, 264629, 18108374, 18108376, 55810764, 65274791, 35695855, 264631, 264634, 264635, 60431850, 264636, 264638, 60170394, 264639, 83373044, 56526486, 87168518, 60432113, 22278000, 22278002, 264564, 264566
1729	91227948 (3457, 3458)	Novel Protein sim. GBank gij854065jemb CAA58337  - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264906, 264907, 264908, 264511, 264555, 83373044, 264596, 264566
1730	85483474 (3459, 3460)			UNCLASSIFIED	29331822, 29331825, 29331828, 264907, 264908, 264909, 265011, 264764, 264629
1731	88266068 (3461, 3462)	Novel Protein sim. GBank gij631600jpir S47094 - hypothetical protein - rabbit		UNCLASSIFIED	52646842, 264907, 264909, 56182435, 55811386, 87168559, 265018, 265019, 264760, 52644229, 55811576
1732	91218878 (3463, 3464)	Novel Protein sim. GBank gij4240231dbj BAA74894.1  - (AB020678) KIAA0871 protein [Homo sapiens]		struct	56182575, 29331822, 29331824, 29331827, 66712502, 264591, 33657402, 60433356, 265019, 21908768, 21908769, 35695917, 285020, 265021, 264836, 59182323
1733	87617178 (3465, 3466)	Novel Protein sim. GBank gij1575756 (U70674) - m-Numb [Mus musculus]	Contains protein domain (PF00640) - synthase Phosphotyrosine interaction domain (PTB/PTD)		264907, 264910, 33657402, 265010, 264681, 264683, 264684, 264686, 264769, 264691, 264692, 264693, 264628, 264636, 264556
1734	87795261 (3467, 3468)				264693

1735	88318638 (3469, 3470)	Novel Protein sim. GBank gi 4836807 gb AAD30566.1 AF14679 - (AF14679) PFT27 [Mus musculus]	Contains protein domain (PF01169) - Uncharacterized protein family UPF0016	264488, 18108394, 18108398, 22278998, 60432049, 264259, 29331822, 264908, 265008, 265007, 265008, 265009, 264591, 33657402, 265010, 265011, 87168559, 265017, 18108351, 264682, 18108354, 264769, 264689, 21808765, 21808768, 21908769, 29148629, 29148784, 265021, 265022, 52644150, 18108384, 18108385, 33657109, 18108370, 18108374, 18108380, 18108385, 87168518, 264583, 18108390, 264480, 264259, 68714117, 68712502, 56182435, 265008, 265008, 264910, 60433356, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264584
1736	95362884 (3471, 3472)	Novel Protein sim. GBank gi 4885647 ref NP_005472.1 pTRAP - thyroid hormone receptor-associated protein complex component	UNCLASSIFIED	264480, 264259, 68714117, 68712502, 56182435, 265008, 265008, 264910, 60433356, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264584
1737	68165549 (3473, 3474)	Novel Protein sim. GBank gi 2143607 ref S68695 - BIK protein - rat	Contains protein domain (PF00168) - Kinase C2 domain	29331822, 28331830, 284591, 265011, 265018, 265019, 22279002
1738	85788811 (3475, 3476)	Novel Protein sim. GBank gi 222594 emb CAA69714  - (Y08460) Mdes protein [Mus musculus]	UNCLASSIFIED	264808, 264809, 265008, 264910, 264566
1739	87328576 (3477, 3478)		Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	35896052, 264603, 264557
1740	83592839 (3479, 3480)	Novel Protein sim. GBank gi 4809 emb CAA44309  - (X82452) YCR801 [Saccharomyces cerevisiae]	UNCLASSIFIED	264604, 21906764, 18108364, 264629, 35895855, 264636
1741	95010100 (3481, 3482)	Novel Protein sim. GBank gi 4836898 gb AAD31695.1 AF13042 - (AF13042) serine protease-like protein isoform [Homo sapiens]	Contains protein domain (PF01363) - FYVE zinc finger	60432229, 29331827, 264509, 265009, 60432229, 264759, 265017, 264767, 264688, 264689, 21908768, 265020, 33657109
1742	85788814 (3483, 3484)	Novel Protein sim. GBank gi 4505193 ref NP_003667.1 pMLD  - membrane fatty acid (lipid) desaturase	UNCLASSIFIED	264806, 264910, 264758, 265011, 264631, 264638, 264566
1743	88866475 (3485, 3486)	Novel Protein sim. GBank		265017, 265020, 264692
1744	91224003 (3487, 3488)	gi 728832 sp P39188 ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII	UNCLASSIFIED	22278998, 264508, 264907
1745	20290075 (3489, 3490)	Novel Protein sim. GBank	UNCLASSIFIED	264559
1746	94328110 (3481, 3492)	gi 731758 sp P38873 YHY6_YEAST - HYPOTHETICAL 175.9 KD PROTEIN IN GND1-IK1 INTERGENIC REGION	Contains protein domain (PF00400) - WD domain, G-beta repeat	65274791, 264639, 264559
1747	94324333 (3493, 3494)	Novel Protein sim. GBank gi 1658503 U75467  - Alu [Drosophila melanogaster]	transcript factor	52546842, 26331824, 26331825, 52644045, 56182435, 265007, 52648317, 87168474, 265018, 18108351, 264369, 264769, 264689, 35895917, 60170615, 264691, 33657023, 65274791, 35895855, 60170394, 22279000, 264482



1748	88003580 (3495, 3498)	Novel Protein sim. GBank gi 450451 ref NP_001530.1 pHSJ2 - heat shock protein, DNAJ-like 2	Contains protein domain (PF00684) - eph DnaJ central domain (4 repeats)	264489, 56182575, 29331824, 56182435, 264112, 265007, 265019, 264764, 21906768, 265020, 264691, 55811576, 264635, 264555, 264556, 264557, 264559 264106
1749	83363091 (3497, 3498)	Novel Protein sim. GBank gi 5650780 gb AAD45948.1 AF151968 - (AF151968) RGS protein RGS-17 (Gallus gallus)	Contains protein domain (PF00615) - oncogene	33657402, 264288, 52644150, 263974, 83373044
1750	94321664 (3499, 3500)	Novel Protein sim. GBank gi 4988894 gb AAC28444.2  - (AF065164) hyperpolarization-activated, cyclic nucleotide- gated channel 2 (Homo sapiens)		
1751	83373058 (3501, 3502)	Novel Protein sim. GBank gi 276016 gb BAA24184  - (AB010054) outer arm dynein light chain 2 (Anthodidaris crassispina)	Contains protein domain (PF00560) - Leucine Rich Repeat	265010, 264369
1752	86458530 (3503, 3504)	Novel Protein sim. GBank gi 3915482 sp P74346 YG29_S1 - HYPOTHETICAL 38.0 KD PROTEIN SLR1629	Contains protein domain (PF00849) - RNA pseudouridylation synthase	264510, 264593, 264682, 21906765, 18108370
1753	94235159 (3505, 3508)	Novel Protein sim. GBank gi 2852836 AF007155  - unknown (Homo sapiens)	Contains protein domain (PF01553) - phosphatase Acyltransferase	56994075, 22278996, 264908, 60170831, 264682, 264754, 264369, 264288, 264685, 264687, 21906765, 264692, 264693, 65274620, 65274791, 35685855, 264637, 264564
1754	88093323 (3507, 3508)	Novel Protein sim. GBank gi 731421 sp P39981 YE4 - HYPOTHETICAL 53.3 KD PROTEIN IN HXT8-CAN1 INTERGENIC REGION	transport	264488, 35698286, 264509, 264906, 264907, 264908, 264909, 264511, 264910, 264991, 33657402, 264594, 264757, 264758, 264600, 264604, 264762, 264763, 264685, 264768, 264691, 264628, 35698423, 264632, 264634, 264637, 264638, 263981, 264558, 264639, 264563, 264564, 264565, 264566, 264567, 264686
1755	78470282 (3509, 3510)	Novel Protein sim. GBank gi 1178422 U43184  - rhophilin [Mus musculus]	UNCLASSIFIED	
1756	92862614 (3511, 3512)	Novel Protein sim. GBank gi 4432860 gb AAD20708  - (AC006300) putative glucose-induced repressor protein [Arabidopsis thaliana]		5264842, 22278994, 22278995, 56984075, 22278996, 22278997, 264259, 29331822, 60432288, 29331827, 33658970, 265008, 265009, 60432229, 60433356, 60433438, 33109954, 21906754, 265017, 265018, 265019, 264448, 264369, 264288, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 265022, 264692, 27486282, 27486284, 18108378, 20281152, 264558, 18108388, 87188518, 60432113, 22279000, 22279002, 264482
1757	85357380 (3513, 3514)	Novel Protein sim. GBank gi 5441615 emb CAB46856.1  - (AJ388557) zinc finger protein [Canis familiaris]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	22278997, 264259, 60432289, 29331827, 264908, 52644045, 265008, 264593, 265019, 264766, 21906768, 65274620, 18108385, 60432113, 264586, 264487

1758	87612971 (3515, 3518)	Novel Protein sim. GBank gij3881040[embjCAAT6403] - (AL021497) predicted using GeneFinder [Caenorhabditis elegans]	UNCLASSIFIED	22278994, 22278999, 29331822, 29331824, 29331825, 33656970, 264508, 265008, 285007, 265009, 264591, 33657402, 33109854, 87188474, 264600, 265017, 265018, 21906769, 265020, 265021, 33657023, 33657109, 264629, 18108374, 35695855, 264632, 52644332, 22278002, 264563
1759	36994372 (3517, 3518)	Novel Protein sim. GBank gij5262748[embjCABA5886.1] - (AJ133120) Proline rich synapse associated protein 2 [Rattus norvegicus]	UNCLASSIFIED	264759
1760	87329716 (3518, 3520)		UNCLASSIFIED	56182575, 60432048, 35696052, 264805, 264906, 264907, 264908, 264909, 265008, 285009, 264910, 60432229, 264592, 264595, 55812038, 264758, 264762, 18108351, 264764, 264766, 264768, 264769, 21906765, 55811957, 35695917, 264690, 264692, 264628, 264629, 55811576, 35696423, 264632, 264634, 264638, 264557, 264639, 80432113, 264565, 264486
1761	87409586 (3521, 3522)	Novel Protein sim. GBank gij127749[sp P10568 MYSC_ACACA - MYOSIN IC HEAVY CHAIN	UNCLASSIFIED	29331822, 264910, 264685, 264686
1762	95319887 (3523, 3524)	Novel Protein sim. GBank gij3169158 (AC004770) - BC269730_2 [Homo sapiens]	Contains protein domain (PF00618) - Guanidine nucleotide exchange factor for Ras-like GTPases; N-terminal motif	22278999, 284259, 284905, 284907, 60170831, 265010, 265011, 265017, 264448, 21908785, 21908786, 21908787, 21908788, 265021, 264690, 33657109, 18108374, 264558, 60170394
1763	91224013 (3525, 3526)	Novel Protein sim. GBank gij4809028[gb AAD30082.1] - (AF132856) suppressor of G2 allele of skp1 homolog [Homo sapiens]	Contains protein domain (PF00173) - desaturase	56181866, 29331825, 35696052, 264905, 264906, 264909, 264763, 264682, 264769, 35695917, 265022, 33657023, 18108374, 35696423, 264634
1764	87757697 (3527, 3528)	Novel Protein sim. GBank gij1360569[pirj CGHU1V - collagen alpha 1(V) chain precursor - human	Contains protein domain (PF01410) - collagen	22278998, 264259, 35696052, 29331828, 56182435, 265008, 265017, 265018, 264448, 284288, 21908788, 21908787, 29148627, 35695917, 264691, 33657023, 60432113, 22279002
1765	91230091 (3529, 3530)	Novel Protein sim. GBank gij486806[pirj S35503 - finger protein neutralized - fruit fly (Drosophila melanogaster)]	UNCLASSIFIED	264488, 18108394, 22278995, 22278997, 22278999, 264259, 264104, 264508, 264905, 264907, 264511, 265007, 265008, 264810, 285009, 21908754, 265010, 265017, 264603, 285018, 285019, 18108351, 264682, 264448, 264369, 264288, 264766, 18108359, 21908766, 21908767, 29148627, 29148629, 35695917, 265020, 265021, 264692, 264628, 264629, 18108374, 263976, 264636, 83373044, 22279000, 22279002

1766	95081201 (3531, 3532)	Novel Protein sim. GBank gi 2499087 sp D09332 UGG_DROME - UDP- GLUCOSE:GLYCOPROTEIN GLUCOSYLTRANSFERASE PRECURSOR [DUGT]		glycoprotein	52845158, 87188559, 60170615, 33657023, 264893, 33657109, 27488261, 264555, 83373044
1767	87755998 (3533, 3534)	Novel Protein sim. GBank gi 417643 emb CAA18263.1  - (AL022238) dJ1042K10.4 (novel protein) [Homo sapiens]		UNCLASSIFIED	265017, 265019, 264686, 264768, 265020, 264692
1768	80253216 (3535, 3536)			UNCLASSIFIED	29331824, 29331825, 264591, 56182323 264563
1769	87388988 (3537, 3538)			UNCLASSIFIED	264488, 264768, 264769, 56182375, 55811957, 264690, 264691, 35696052, 264905, 264509, 264906, 264907, 264628, 264908, 264909, 264910, 264634, 264635, 264636, 264556, 264757, 264758, 55812038, 65274444, 264760, 264563, 264762, 264764, 264684, 264768
1770	95413144 (3539, 3540)	Novel Protein sim. GBank gi 15204 sp P00736 C1R_HUMAN - COMPLEMENT C1R COMPONENT PRECURSOR	Contains protein domain (PF00089) - Trypsin	complement	
1771	94233542 (3541, 3542)	Novel Protein sim. GBank gi 391418 lsp P56558 OGT1_RAT - UDP-N- ACETYLGLUCOSAMINE-PEPTIDE N- ACETYLGLUCOSAMINYLTTRANSFERASE 110 KD SUBUNIT (O-GLCNAC TRANSFERASE P110 SUBUNIT)	Contains protein domain (PF00515) - TPR Domain	transferase	22278998, 26331828, 33109954, 265018, 265019, 264764, 21906768, 265020, 265021, 264558
1772	87643510 (3543, 3544)	Novel Protein sim. GBank gi 4959442 gb AA034351.1 AF12136 - (AF121360) DNZDHH/CNEW1 zinc finger protein 11 [Drosophila melanogaster]	Contains protein domain (PF01529) - DHH zinc finger domain		
1773	84118824 (3545, 3546)	Novel Protein sim. GBank gi 3978464 (AF085693) - G protein-coupled receptor kinase-associated ADP ribosylation factor GTPase-activating protein [Rattus norvegicus]	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	potassium_channel	65274572, 56182575, 22278996, 35696286, 22278999, 264259, 29331824, 60424269, 29331825, 60432289, 35696052, 264108, 264509, 264908, 264907, 29331830, 264908, 52644045, 264511, 265008, 265007, 265008, 60170831, 60433438, 264758, 55811388, 87188559, 265017, 264604, 265019, 55811150, 264288, 56181562, 264689, 21906768, 21906767, 21906768, 21906769, 55811957, 265020, 265022, 52644150, 264691, 33657023, 264692, 264693, 60431528, 35696423, 35695855, 264636, 56182323, 18108387, 56526488, 22278000, 22279002, 264563, 264584, 264585, 264586, 264587
1774	84232573 (3547, 3548)	Novel Protein sim. GBank gi 2495699 sp Q15034 Y032_HUMAN - HYPOTHETICAL PROTEIN KIAA0032	Contains protein domain (PF00415) - Regulator of chromosome condensation (RCC1)	UNCLASSIFIED	65274572, 56182575, 35696052, 55812038, 33109954, 21906754, 265017, 21906767, 21906768, 21906769, 265020, 264691, 264636, 56182323, 22279002

1775	9359330 (3549, 3550)	Novel Protein sim. GBank gi 1468169 dbj BAA09487.1 - (D50928) The KIAA0139 gene product is novel. [Homo sapiens]	UNCLASSIFIED	65274572, 56182575, 56994075, 35696286, 22278997, 29331822, 29331826, 60432289, 29331828, 35696052, 29331830, 68712502, 264828, 56182435, 264511, 265007, 265009, 60170831, 60432229, 60433438, 55812036, 21908754, 85658542, 87188559, 264601, 265017, 265018, 265019, 264762, 264448, 264288, 264688, 21906765, 21906766, 21906767, 21906768, 265020, 265021, 265022, 264691, 33657023, 264692, 33657109, 27486261, 33657349, 18108370, 18108377, 35696423, 55811576, 35695855, 264632, 264634, 264636, 264639, 56182323, 83373044, 58526486, 87188518, 60432113, 22279000, 22279002, 264482, 264488, 264910
1776	94133756 (3551, 3552)	Novel Protein sim. GBank gi 4589676 dbj BAA76857.1 - (AB023230) KIAA1013 protein [Homo sapiens]		
1777	87447171 (3553, 3554)	Novel Protein sim. GBank gi 3219339 sp P87115 YDK9 SCHPO - HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME 1	nuc_rept	58994075, 29331826, 265008, 87188474, 265017, 265018, 264761, 55811150, 264764, 56181562, 264689, 21906765, 21906768, 21906769, 35695917, 264690, 33657023, 35695763, 60431528, 35696423, 55811576, 35695855, 22279000, 22279002, 264564
1778	94851624 (3555, 3556)	Novel Protein sim. GBank gi 3875648 emb CAA91454.1 - (Z66561) Similarity to Human rab13 protein (PIR Acc. No. A49647). Contains the ATP/GTP-binding site motif (PROSITE PS00017); cDNA EST EMBL: M89412 comes from this gene; cDNA EST yk212g9.3 comes from this gene; cDNA EST yk212g9...	UNCLASSIFIED	29331826, 29331827, 35696052, 264512, 265007, 265009, 265017, 265018, 264762, 18108351, 264769, 21906765, 21906766, 21908767, 21908768, 21908769, 265020, 265021, 264691, 264693, 18108370, 264556, 83373044
1779	94133758 (3557, 3558)	Novel Protein sim. GBank gi 4589676 dbj BAA76857.1 - (AB023230) KIAA1013 protein [Homo sapiens]	UNCLASSIFIED	29148627, 35696286, 29147620, 265006, 265007, 265008, 18108385, 65274727, 264482, 264369, 264766
1780	87023497 (3559, 3560)		Contains protein domain (PF00807) - Apidaedn	264107, 33657109, 56526486
1781	84047477 (3561, 3562)		UNCLASSIFIED	264508, 264908, 264639
1782	88094607 (3563, 3564)	Novel Protein sim. GBank gi 729225 sp P41237 CTXN_RAT - CORTEXIN		264259, 29331822, 264508, 264905, 264906, 264907, 264908, 265007, 265009, 264910, 264591, 264758, 264764, 264288, 264768, 264769, 264635, 264636, 264637, 264639, 264563
1783	85717805 (3565, 3566)	Novel Protein sim. GBank gi 225754 dbj BAA21436.1 - (AB004538) protein arginine N-methyltransferase [Schizosaccharomyces pombe]	interferon	264788

1784	95197083 (3567, 3568)	Novel Protein sim. GBank gij1755049 (U55042) - myosin X [Bos taurus]	Contains protein domain (PF00169) - struct PH domain	3569286, 264259, 35696052, 264508, 264905, 264906, 264907, 68712502, 264908, 264909, 265007, 265008, 265009, 264910, 264591, 264694, 264757, 264758, 264759, 265010, 265011, 264601, 264602, 264604, 264605, 18108351, 264762, 264763, 264764, 264389, 264768, 264687, 264768, 264688, 21908768, 35695917, 264690, 264691, 264692, 264693, 264628, 18108374, 35696423, 264631, 264632, 264635, 264637, 264638, 264639, 18108385, 22279000, 22279002, 264563, 264566, 264486
1785	95357475 (3569, 3570)	Novel Protein sim. GBank gij4589552(bj)BAA76798.1] - (AB023171) KIAA0954 protein [Homo sapiens]	UNCLASSIFIED	65274572, 264259, 29331822, 29331824, 29331825, 29331827, 29331828, 264905, 264906, 264908, 68712502, 56182435, 264511, 265007, 60433356, 55811150, 264683, 264369, 264687, 52644229, 21908767, 52644150, 33657023, 65274620, 33657182, 65274791, 35695855, 264555, 65274727, 22279002
1786	85296465 (3571, 3572)	Novel Protein sim. GBank gij117788(isp)P26770(CY44_RAT ADENYLATE CYCLASE, TYPE IV (ATP PYROPHOSPHATE-LYASE) (ADENYL CYCLASE)	UNCLASSIFIED	264908, 35696423, 264636
1787	87434784 (3573, 3574)	Novel Protein sim. GBank gij3877175(emb)CAA90338.1] - (Z50028) cDNA EST yk321h8.5 comes from this gene; cDNA EST EMBL:D68896 comes from this gene; cDNA EST yk395f9.5 comes from this gene [Caenorhabditis elegans]	UNCLASSIFIED	264488, 264905, 264908, 264909, 264595, 264764, 264766, 264692, 60431528, 264629, 264636, 264584, 264566
1788	91228779 (3575, 3576)		UNCLASSIFIED	264488, 63373044
1789	86094529 (3577, 3578)	Novel Protein sim. GBank gij2088669 (AF003130) - F55A12.9 gene product [Caenorhabditis elegans]	UNCLASSIFIED	264488, 29331828, 264909, 18108351, 264288, 265021, 264555, 264636
1790	82469734 (3579, 3580)			35696052, 264803, 264906, 264807, 264908, 264909, 265008, 264910, 264738, 265011, 265019, 264764, 264768, 264769, 264628, 264635

1781	95197259 (3581, 3582)	Novel Protein sim. GBank gij2114321[dbjBAA20037] - (D88733) membrane glycoprotein [Equine herpesvirus 1]	Contains protein domain (PF00047) - glycoprotein immunoglobulin domain	264488, 264686, 264687, 264768, 18108394, 264769, 18108397, 264259, 264681, 264692, 33857023, 264693, 264698, 264905, 264906, 264628, 264907, 264628, 264908, 264909, 264510, 265008, 264511, 265008, 264630, 265009, 264631, 264910, 264632, 264634, 264635, 264555, 264636, 264592, 264637, 264583, 264638, 18108381, 264639, 264759, 265010, 265011, 264602, 22279000, 264604, 264760, 264564, 264681, 264762, 264565, 264763, 264683, 264566, 264764, 264288, 264684, 264587, 18108354, 18108391, 264685, 264766
1782	87792680 (3583, 3584)	Novel Protein sim. GBank gij4337106[gb AAD16082] - (AF128756) BAT4 [Homo sapiens]	Contains protein domain (PF01585) - G-patch domain	22278997, 264259, 264508, 265007, 33857402, 87188559, 264369, 33857023, 35695855, 20281071, 264359, 18108387, 87188518
1783	95337877 (3585, 3586)	Novel Protein sim. GBank gij5579331[gb AAD45504.1] (AF145732) endoplasmic reticulum alpha-mannosidase I [Homo sapiens]	Contains protein domain (PF01532) - Glycosyl hydrolase family 47	85274572, 22278995, 22278996, 22278997, 22278999, 264093, 264259, 29331824, 66714117, 60432289, 29331827, 29331828, 264103, 264105, 29331830, 255007, 264910, 265009, 60170831, 60433356, 21906754, 265010, 265017, 265019, 264681, 264682, 264288, 52644228, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 265022, 60170615, 52644150, 33857023, 33857109, 18108370, 18108374, 85274791, 20281071, 60432113, 22279000, 264482, 264564
1784	87759806 (3587, 3588)	Novel Protein sim. GBank gij4914604[emb CAB3677.1] - (AL050369) hypothetical protein [Homo sapiens]	Contains protein domain (PF01788) - Putative snoRNA binding domain	18108394, 22278995, 22278999, 264259, 29331822, 29331824, 29331825, 29146498, 29146499, 264508, 264905, 52644045, 264112, 265006, 265008, 264910, 60433356, 264757, 55612038, 87168474, 265011, 265017, 18108351, 264763, 264448, 264683, 264369, 21906765, 21906766, 21906767, 21906769, 29148784, 35695817, 60170615, 33857023, 284628, 18108374, 18108376, 35696423, 35695855, 264556, 264557, 264538, 264558, 18108385, 264584, 264632, 264635, 264636, 264595, 264596, 264907, 264566, 264909
1785	79747856 (3589, 3590)		UNCLASSIFIED	
1786	86599486 (3591, 3592)	Novel Protein sim. GBank gij585084[sp Q07803 IEFGM_RAT - ELONGATION FACTOR G, MITOCHONDRIAL PRECURSOR (MEF-G)]	glycoprotein	264488, 264907, 264909, 264594, 264595, 264766, 264687, 21906765, 21906767, 264628, 264630, 264559

1797	91223218 (3593, 3594)	Novel Protein sim. GBank gi 1842111 (UB7586) - decoy (Arabidopsis thaliana)		ribosomalprol	22278996, 22278997, 22278998, 22278999, 29331822, 264910, 60170831, 21908754, 52644229, 21908785, 21908788, 21908789, 35695917, 265022, 52644150, 264691, 33657023, 263967, 33657109, 22279000
1798	91221276 (3595, 3596)	Novel Protein sim. GBank gi 2832905(dbj BAA24808.1) - (D89340) dipeptidyl peptidase III [Rattus norvegicus]		peptidase	22278994, 56984075, 22278997, 22278998, 22278999, 264259, 29331826, 60432289, 29331828, 33858970, 265008, 60432229, 264757, 60433438, 21908754, 33857084, 87168559, 265017, 18108351, 264682, 264448, 264288, 21908785, 21908788, 21908787, 21908788, 21908789, 35695917, 285020, 265021, 33857023, 33857182, 27488281, 27488285, 33657348, 283973, 18108374, 55811578, 35695855, 18108385, 87168518, 22279000, 264486
1799	86321713 (3597, 3598)	Novel Protein sim. GBank gi 568954 (dbj BAA83054.1) - (AB029025) KIAA1102 protein [Homo sapiens]		eph	284908, 21908754, 21908787, 21908789, 285020, 33657023, 284692, 264693, 264404, 22279000
1800	87080116 (3598, 3600)	Novel Protein sim. GBank gi 4680678(gb AAD27729.1 AF132954) CGI-20 protein [Homo sapiens]		UNCLASSIFIED	284691, 284556, 284566
1801	95060723 (3601, 3602)	Novel Protein sim. GBank gi 4680678(gb AAD27729.1 AF132954) CGI-20 protein [Homo sapiens]			52844045, 265007, 264632
1802	87771012 (3603, 3604)	Novel Protein sim. GBank gi 134920(sp P21997(SSGP_VOLCA - SULFATED SURFACE GLYCOPROTEIN 185 (SSG 185)			35695286, 66714117, 264508, 264509, 56182435, 264512, 18108351, 284688, 55811957, 264692, 55811576, 35695855, 284488
1803	95060725 (3605, 3608)	Novel Protein sim. GBank gi 4680678(gb AAD27729.1 AF132954) CGI-20 protein [Homo sapiens]			284688, 284488, 284687, 284489, 284768, 264769, 264689, 21908788, 35696288, 35695917, 264259, 264691, 264692, 264693, 20281099, 18108384, 35696052, 264508, 264509, 264905, 264906, 18108370, 264628, 284907, 68712502, 264908, 264909, 18108374, 18108376, 35696423, 35695855, 264510, 264511, 265008, 265007, 264512, 265008, 264910, 264831, 264832, 264634, 264635, 264591, 264836, 264837, 264592, 264638, 264593, 264839, 264594, 83373044, 284758, 264586, 18108385, 18108387, 265011, 264760, 264583, 18108351, 264762, 264584, 264448, 264585, 264763, 264683, 284764, 264586, 264288, 264486, 264567, 264765, 264766

1804	87770203 (3607, 3608)	Novel Protein sim. GBank gij3879914[embjCAA98538.1] - (Z74043) predicted using GeneFinder. cDNA EST EMBL: C13850 comes from this gene; cDNA EST EMBL: C11575 comes from this gene; cDNA EST yk3434.5 comes from this gene [Caenorhabditis elegans]			52646365, 22278997, 22278999, 264905, 264908, 264909, 264910, 21808754, 264768, 21908765, 21908768, 35695817, 265020, 265022, 264691, 264637, 264639, 22279000, 264584, 264568
1805	95330375 (3608, 3610)	Novel Protein sim. GBank gij545364[refjNP_006461.1]pEBBP - estrogen-responsive B box protein			29331824, 29331825, 29331826, 29331827, 29331828, 87168559, 264288, 264687, 52644229, 35698423, 264636, 60432113
1806	94133782 (3611, 3612)	Novel Protein sim. GBank gij4589876[jb]BAA76857.1] - (AB023230) KIAA1013 protein [Homo sapiens]	struct		264094, 264105, 264908, 35698423, 265008, 265007, 265008, 264555, 264592, 265011, 265018, 264369
1807	86943032 (3613, 3614)				29331824, 264908, 264910, 33657023, 263978
1808	87642711 (3615, 3616)	Novel Protein sim. GBank gij4884079[embjCAB43235.1] - (AL050008) hypothetical protein [Homo sapiens]	UNCLASSIFIED		264488, 35696286, 66714117, 35698052, 66712502, 264592, 60433438, 52644286, 265010, 264683, 264369, 264689, 55811957, 35695917, 33657108, 35695783, 55810764, 18108379, 35696423, 35695855, 56182323, 264583, 264584, 264487
1809	95321468 (3617, 3618)	Novel Protein sim. GBank gij1916827 (U87865) - putative G protein [Mus musculus]	UNCLASSIFIED		264594, 55811150, 264686, 29148629, 29148784, 264690, 264629, 18108374, 264556, 264557, 264558
1810	88096316 (3619, 3620)	Novel Protein sim. GBank gij1352944[sp]P47178YJ9P_YEAST - HYPOTHETICAL 118.4 KD PROTEIN IN BAT2-DAL5 INTERGENIC REGION PRECURSOR	UNCLASSIFIED		264488, 35696052, 264905, 264906, 264907, 264908, 264909, 264511, 265009, 264910, 264592, 264593, 264594, 33657402, 264757, 264595, 264758, 264596, 264759, 264600, 264601, 264762, 264683, 264764, 264288, 264684, 264768, 264787, 264688, 264768, 264687, 264769, 264689, 265021, 264690, 264691, 264693, 264628, 264628, 18108374, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 264639, 264583, 264586, 264486, 264587
1811	88086272 (3621, 3622)	Novel Protein sim. GBank gij2134984[pir]I37275 - death-associated protein kinase (EC 2.7.1.-) - human	kinase Ank repeat	Contains protein domain (PF00023) - Ank repeat	264488, 264239, 264508, 264509, 264905, 264906, 264907, 56182435, 264511, 264512, 264910, 264758, 265011, 264600, 264604, 18108354, 264766, 264686, 264769, 264534, 60170515, 33657023, 264629, 264831, 264639, 264583, 264482, 264483
1812	79245772 (3623, 3624)				29331822, 29331824, 265019, 18108351, 21808769



1813	86090972 (3625, 3626)	Novel Protein sim. GBank gij5051638jg/AAD38326.1 AF07372.2 - (AF073727) EH domain-binding mitotic phosphoprotein [Homo sapiens]	Contains protein domain (PF01417) - ENTH domain	UNCLASSIFIED	55182575, 284259, 29331824, 66714117, 29331828, 35896052, 284509, 264803, 264906, 284907, 284908, 66712502, 264909, 265007, 264910, 264591, 264593, 55812038, 265011, 265018, 264760, 264682, 264784, 264683, 264369, 264766, 264788, 264769, 21906766, 21906768, 264691, 264693, 18108374, 35895855, 264634, 264635, 264637, 264639, 264559, 22279000, 22279002, 264568
1814	85178047 (3627, 3628)	Novel Protein sim. GBank gj13643608 (AC005395) - hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	264488, 35896286, 22278998, 264092, 264094, 264258, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 264105, 264107, 52844045, 56182435, 265009, 60432229, 60433356, 87168474, 87168559, 264389, 264288, 21906765, 35895917, 265021, 265022, 33657023, 33657109, 18108374, 35898423, 264638, 56526485, 264482
1815	85298473 (3629, 3630)	Novel Protein sim. GBank gj117788 sp P26770 CYA4_RAT ADENYLATE CYCLASE, TYPE IV (ATP PYROPHOSPHATE-LYASE) (ADENYL CYCLASE)	sinud		22278999, 264508, 264509, 264907, 264908, 264910, 265011, 264760, 264766, 264634, 264636
1816	83738845 (3631, 3632)	Novel Protein sim. GBank gij1176623 sp P41846 YO98_CAEEL - HYPOTHETICAL 93.9 KD PROTEIN T20812.6 IN CHROMOSOME III	UNCLASSIFIED		18108394, 18108397, 284508, 264907, 264908, 264909, 285009, 264591, 265011, 265017, 264687, 264689, 265022, 264691, 18108392, 18108398, 18108370, 18108374, 18108379, 264635, 264537, 264564, 264587
1817	86095268 (3633, 3634)	Novel Protein sim. GBank gj13766377 emb CAA21428  - (AL031907) hypothetical protein [Schizosaccharomyces pombe]	Contains protein domain (PF00400) - WD domain, G-beta repeat	histone	264488, 22278997, 22278999, 60432049, 29331822, 29331824, 60432289, 52844045, 60170831, 265017, 265018, 265019, 18108351, 264682, 52844229, 21906765, 21906767, 21906768, 52844150, 33657023, 33657109, 27486262, 18108370, 18108374, 60170394, 56182323, 22279002
1818	85806775 (3635, 3636)	Novel Protein sim. GBank gj1387912 emb CAA94370  - (Z70310) predicted using GeneFinder. Similarity to Mouse ankyrin (Pir Acc. No. S3771); cDNA EST EMBL.T01923 comes from this gene; cDNA EST EMBL.D32335 comes from this gene; cDNA EST EMBL.D32723 comes from this gene; cDNA EST E...	Contains protein domain (PF00023) - Ank repeat	transcript factor	35896286, 60433356, 264758, 264369, 264686, 21906769, 264693, 264632
1819	87759372 (3637, 3638)	Novel Protein sim. GBank gij5031865 ref NP_005771.1 pLHFP - lipoma HMGIC fusion partner	UNCLASSIFIED		22278995, 22278997, 22278999, 264259, 29331826, 264908, 265007, 265008, 265009, 60432229, 265017, 265018, 265019, 264448, 264288, 21906768, 21906769, 265020, 18108381, 18108384, 22279000, 22279002, 264587

1820	87769455 (3639, 3640)				264905, 264907, 264584
1821	80431510 (3641, 3642)				264907, 264768, 263978
1822	81221523 (3643, 3644)	Novel Protein sim. GBank gi 4884130 emb CAB43272.1  - (AL050101) hypothetical protein (Homo sapiens)			22278995, 56994075, 22278996, 22278997, 22278998, 264259, 29331824, 29331825, 29331826, 35695052, 29331828, 264908, 29331830, 60170831, 264591, 264593, 60433356, 264596, 265017, 265019, 18108351, 264753, 264683, 21906765, 21906767, 21906768, 21906769, 35695817, 265020, 265021, 33657023, 18108364, 18108370, 35695855, 22279000, 22279002, 264488, 264259, 264511, 264286, 264766, 264693, 35696423, 264634, 18108385, 264488
1823	85522330 (3645, 3646)			UNCLASSIFIED	
1824	86612025 (3647, 3648)	Novel Protein sim. GBank gi 477072 pr IA48018 - mucin 7 precursor, salivary - human	Contains protein domain (PF00047) - Immunoglobulin domain	UNCLASSIFIED	264907, 264908, 264809, 264511, 264631, 264634, 264635, 264637, 264638, 264639, 264758, 264568
1825	87430125 (3649, 3650)	Novel Protein sim. GBank gi 3036603 emb CAA18493  - (AL022373) hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	60432049, 264910, 264487
1826	91723612 (3651, 3652)	Novel Protein sim. GBank gi 468065 gb AAD27732.1 AF13295 - (AF132957) CGI-23 protein (Homo sapiens)		ATPase associated	52644507, 52645156, 52646842, 22278994, 22278996, 56994075, 264259, 60432049, 52645080, 35696052, 66712502, 52644045, 265008, 265009, 60432229, 60433366, 60433438, 52646317, 52644286, 265011, 87168559, 264448, 264288, 264369, 264688, 52644229, 264689, 21906765, 21906768, 265020, 60170615, 52644150, 33657023, 27486262, 27486264, 27486265, 35695763, 35696423, 35695855, 83373044, 87168518, 264404, 22279002
1827	81647212 (3653, 3654)				264758

1826	85074017 (3655, 3656)	Novel Protein sim. GBank gll4503571[ref(NP_001419.1)pEN01 - enolase 1, (alpha)]	Contains protein domain (PF00113) - End-ase	onco gene	264488, 52645842, 56182575, 22278996, 35959286, 22278997, 22278999, 264091, 264093, 60432048, 264259, 29331822, 29331824, 66714117, 29331825, 60432289, 29331826, 29331827, 29331828, 264105, 264508, 264907, 68712502, 52644045, 56182435, 265006, 284511, 264512, 265007, 265008, 265009, 60170831, 60432228, 264593, 60433356, 60433438, 264758, 33109954, 21908754, 87168474, 265010, 265011, 87168559, 265017, 265019, 264761, 264762, 264448, 264764, 264683, 264288, 264369, 18108355, 264768, 18108357, 18108358, 264688, 264769, 264689, 21908768, 21908769, 35695917, 265021, 60170815, 33657023, 33657349, 263972, 55811576, 35695855, 264635, 264555, 264556, 264638, 284557, 87188518, 22279000, 22279002, 264563, 264482, 264555, 264484, 264567
1829	80187720 (3657, 3658)				264508, 264634, 264509, 264482, 29331827, 264908, 265009, 264910
1830	94312942 (3659, 3660)	Novel Protein sim. GBank gll2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma- associated herpesvirus]	nuclease		52645156, 22278994, 22278995, 35695286, 22278996, 22278997, 22278998, 22278999, 29331822, 29331825, 35696052, 52646317, 52644296, 87168559, 265019, 21908765, 21908766, 21908767, 21908768, 21908769, 35695917, 285021, 33657023, 52645129, 33657109, 33657182, 27486281, 27486282, 35695783, 263974, 35696423, 35695855, 52644332
1831	94138063 (3661, 3662)		UNCLASSIFIED		29331824, 35696052, 29331830, 264595, 264758, 265010, 265019, 265022, 264693, 65274791
1832	84521663 (3663, 3664)	Novel Protein sim. GBank gll1330345 (U58755) - coded for by C. elegans cDNA yk34b1.5; coded for by C. elegans cDNA yk13h10.5; coded for by C. elegans cDNA yk46e8.5; coded for by C. elegans cDNA yk46d5.5; coded for by C. elegans cDNA yk43c2.5; coded for by C. elegans cDNA yk46e8....			264602

1833	95314184 (3665, 3666)	Novel Protein sim. GBank gij5174413[re]NP_006026.1pCDC4 - CDC42-binding protein kinase beta (DMPK-like)	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	22278994, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 60432289, 29331827, 35696052, 29146499, 264508, 264509, 264906, 264907, 68712502, 264908, 52844045, 264809, 264512, 265006, 264591, 264593, 60433356, 21906754, 33657084, 265011, 265017, 264804, 265018, 265019, 264681, 18108351, 264883, 264288, 264885, 264788, 264687, 21906765, 21906766, 21906787, 21906768, 21906769, 29148828, 265020, 265021, 264690, 264692, 33657023, 65274620, 33657182, 27486264, 33657349, 65274791, 264634, 264635, 264556, 264557, 264558, 264559, 18108385, 56528486, 87168518, 60432113, 22279000, 22279002, 264563
1834	80562790 (3667, 3668)				264259, 264907, 264689, 22279000, 22279002
1835	94135718 (3669, 3670)			UNCLASSIFIED	22278998, 29331822, 29331826, 87168474, 264603, 21906788, 263976, 35696855, 83373044
1836	87348450 (3671, 3672)	Novel Protein sim. GBank gij4759286[re]NP_004268.1pUCP4 - uncoupling protein 4	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	29331825, 264908, 265019, 264764, 264688, 21906765, 264635
1837	94234297 (3673, 3674)	Novel Protein sim. GBank gij333400[sp]Q24574IUBPE_DROME - UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 64E (UBIQUITIN THIOLESTERASE 64E) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 64E) (DEUBIQUITINATING ENZYME 64E)	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	22278995, 29146499, 265006, 265008, 265009, 265010, 264683, 21806765, 29148627, 29148628, 265020, 265021, 265022, 65274820, 18108370, 18108374, 264556, 18108385
1838	94324359 (3675, 3676)	Novel Protein sim. GBank gij1362599[pi]A56154 - Abl substrate ena (enabled) - fruit fly (Drosophila melanogaster)	Contains protein domain (PF00568) - WH1 domain		29331822, 265017, 264760, 265020, 83373044
1839	87456508 (3677, 3678)	Novel Protein sim. GBank gij211731[0]enb[CAB09116.1] - (Z95620) hypothetical protein [Schizosaccharomyces pombe]		UNCLASSIFIED	60433438, 264601, 21906765, 21906766, 265021, 33657108, 264556
1840	87391708 (3679, 3680)	Novel Protein sim. GBank gij127960[sp]P23249[MOV10] MOUSE - PROTEIN MOV-10		UNCLASSIFIED	264693
1841	65818445 (3681, 3682)	Novel Protein sim. GBank gij4572464[gb]AAD23834.1[AF12365.3] FEZ1 [Homo sapiens]			56182575, 29331824, 29331826, 60433356, 264764, 264288, 33657023, 263967, 18108370, 18108374, 264631, 264555, 264556, 264639

1842	90892845 (3683, 3684)	Novel Protein sim. GBank gij1326268 (U58728) - C54H2.1 gene product [Caenorhabditis elegans]		UNCLASSIFIED	65274572, 26331822, 26331824, 26331825, 69714117, 29331826, 29331827, 264907, 264909, 52644045, 56182435, 264510, 265006, 265007, 265009, 264910, 60433356, 264757, 60433438, 55812038, 265017, 18108354, 264686, 264769, 33657023, 264693, 18108384, 33657109, 18108368, 264628, 55810764, 56182323, 18108384, 264563, 264564
1843	95292692 (3685, 3686)			UNCLASSIFIED	264488, 56182435, 264769, 29331826, 29331828, 264511, 268008, 265007, 264910, 264631, 264509, 264680, 264636, 264564, 264691, 60432228, 60432048, 264258, 264628, 33657023, 264486, 264908, 264567, 264595, 264768
1844	67444764 (3687, 3688)	Novel Protein sim. GBank gij2486887 [sp] Q9232 [YQ22_CAEEL - HYPOTHETICAL 32.0 KD PROTEIN C99F5.2 IN CHROMOSOME III]		UNCLASSIFIED	264908, 285022, 33657023, 87168518, 22279002
1845	95096673 (3689, 3690)	Novel Protein sim. GBank gij1175494 [sp] Q9819 [YACS_SCHPO - HYPOTHETICAL 45.0 KD PROTEIN C16C9.05 IN CHROMOSOME I]	Contains protein domain (PF00628) - PHD-finger	transcription factor	264259, 29331824, 264907, 264908, 68712502, 264510, 265007, 265008, 55812038, 265018, 21908765, 52644150, 33657109, 264555, 264556, 264557, 56182323, 18108382, 83373044, 18108385, 264564
1846	94287872 (3691, 3692)	Novel Protein sim. GBank gij388108 [emb] CAA21739 [AL032657] similar to EGF-like domain; cDNA EST YK289a12.3 comes from this gene; cDNA EST EMBL:D35388 comes from this gene; cDNA EST YK331n6.5 comes from this gene; cDNA EST YK289a12.5 comes from this gene; cDNA EST YK467g8.5 c...	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	264905, 264908
1847	87821487 (3693, 3694)	Novel Protein sim. GBank gij5059323 [gb] AAD38967.1 [AF151522] hairy and enhancer of split related-1 [Homo sapiens] (AJ245417) G5b protein [Homo sapiens]	Contains protein domain (PF00010) - Helix-loop-helix DNA-binding domain	transcription factor	22278997, 264259, 29331824, 264909, 18108351, 263974, 22279002
1848	86789360 (3695, 3696)	Novel Protein sim. GBank gij5701854 [emb] CAB52191.1 - (AJ245417) G5b protein [Homo sapiens]		tm7	29331825, 29331826, 29331827, 265017, 264683, 264288, 264768, 264768, 21908767, 21908768, 264682, 22279002
1849	94287874 (3697, 3698)	Novel Protein sim. GBank gij4503665 [ref] NP_001989.1 [pFBLN - fibulin 2 precursor (AB023186) KIAA0969 protein [Homo sapiens]	Contains protein domain (PF00008) - EGF-like domain	ATPase-associated	56182575, 265018
1850	86889650 (3699, 3700)	Novel Protein sim. GBank gij4589862 [gb] BAA76813.1 - (AB023186) KIAA0969 protein [Homo sapiens]	Contains protein domain (PF00169) - PH domain	strud	60432049, 264908
1851	95418789 (3701, 3702)	Novel Protein sim. GBank gij220637 [db] BAA01471 - (D10827) zinc finger protein [Mus musculus]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_ma_bind	29331824, 35698052, 264910, 60433438, 264688, 35695917, 265020, 52644150, 65274620, 52644332

1852	95413170 (3703, 3704)	Novel Protein sim. GBank gii5174829[refNP_005090.1]pPIAS - protein inhibitor of activated STAT3		UNCLASSIFIED	58182575, 35898286, 22276998, 22278897, 22278999, 264490, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 60432289, 29331827, 35696052, 52644045, 265007, 264910, 60432229, 60433356, 60433438, 55812038, 65274444, 265018, 265019, 18108351, 264448, 264686, 264687, 21808765, 21808787, 21808789, 265021, 265022, 52644150, 264693, 33657109, 18108370, 18108374, 55811576, 35698555, 58182323, 60432113, 22278002, 264563, 264487, 264768, 52644507, 264769, 21808765, 21808787, 21808788, 22278895, 58994075, 22278998, 52644150, 264259, 264692, 29331822, 29331824, 52645129, 29331827, 33656970, 33657349, 35695763, 264508, 264906, 264628, 284907, 264629, 264909, 35696423, 35695855, 264510, 265006, 264511, 264512, 264630, 265009, 264631, 264910, 264634, 264635, 264637, 264593, 264638, 264639, 33657402, 18108385, 52646317, 52644298, 87168518, 87168559, 284602, 285017, 22279000, 265018, 264760, 264762, 264682, 264448, 264764, 264684, 264567, 264288, 264369, 264768 264592
1853	91222267 (3705, 3706)	Novel Protein sim. GBank gii854065[emb CAA58337] - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	
1854	88038152 (3707, 3708)	Novel Protein sim. GBank gii2072864 (U83569) - putative p150 [Homo sapiens]		nuclease	
1855	91221459 (3709, 3710)	Novel Protein sim. GBank gii4539520[emb CAB39994.1] - (AL035424) dA22D12.1 (novel protein similar to Drosophila Kelch (Ring Canal protein, KEL) and a heterogeneous set of other types of proteins) [Homo sapiens]	Contains protein domain (PF01344) - Kelch motif	- nucl_rept	18108382, 52646385, 65274572, 58182575, 22278999, 58994075, 22278998, 35696286, 22278999, 264259, 60432049, 264908, 264510, 265007, 265008, 265009, 264595, 21808784, 87168474, 265011, 87168559, 264681, 264288, 264768, 18108359, 21808784, 21808788, 28148827, 265020, 265021, 285022, 52644150, 33657023, 33657109, 18108372, 18108374, 18108376, 35696423, 264631, 264636, 18108381, 284482
1856	94231871 (3711, 3712)	Novel Protein sim. GBank gii395497[emb CAA06945] - (AJ006278) acetylglucosaminyltransferase-like protein [Mus musculus]		UNCLASSIFIED	58994075, 264259, 29331828, 264511, 264910, 264758, 264693, 264637, 18108381, 83373044
1857	94324455 (3713, 3714)	Novel Protein sim. GBank gii4322670[gb AAD16120] - (AF094508) dentin phosphoryn [Homo sapiens]		ATPase-associated	22278999, 284259, 284908, 60170831, 264448, 264686, 265020, 265022, 33657109, 60170384, 83373044

1858	87628311 (3715, 3716)	Novel Protein sim. GBank gi 4881903 gb AAD36415.1 AE001788 - (AE001788) ribosomal protein S15 [Thermotoga maritima]	Contains protein domain (PF00312) - ribosomal prot	284757
1859	84407464 (3717, 3718)	Novel Protein sim. GBank gi 4240317 db BAA74937.1  - (AB020721) KIAA0914 protein [Homo sapiens]		22278996, 29331824, 265007, 33109954, 285019, 264369, 21906768, 29148784, 27488261, 52644332, 22279002
1860	17929308 (3719, 3720)	Novel Protein sim. GBank gi 4009522 (AF009731) - connexin 31.1 [Homo sapiens]		265019
1861	86086370 (3721, 3722)	Novel Protein sim. GBank gi 2143637 pir j184505 - calcium- dependent actin-binding protein - rat	Contains protein domain (PF00285) - Citrate synthase	264887, 264259, 29331822, 29331824, 29331825, 265007, 265009, 264391, 33109954, 265010, 265019, 264389, 264288, 264686, 264691, 264693, 27488284, 18108370, 18108374, 263977, 55811576, 58182323, 264639, 22279000, 22279002, 264482
1862	87372923 (3723, 3724)	Novel Protein sim. GBank gi 125493 sp P07313 KMLC_RABIT - MYOSIN LIGHT CHAIN KINASE, SKELETAL MUSCLE (MLCK)	kinase	35696286, 264259, 87168474, 264369, 21906768, 264558, 264563
1863	85775037 (3725, 3726)	Novel Protein sim. GBank gi 3820909 emb CAA09299  - (AJ010842) Def protein [Drosophila melanogaster]	UNCLASSIFIED	264601, 264766, 29148627, 29148628, 264692, 264829, 264835
1864	85547832 (3727, 3728)	Novel Protein sim. GBank gi 4322263 gb AAD15985  - (AF077738) metalloproteinase CPX-1 [Mus musculus]	synthase	22278999, 264259, 264907, 265018, 18108370, 264634, 264635, 264555, 264556, 264638, 18108387
1865	87740827 (3729, 3730)	Novel Protein sim. GBank gi 2485727 sp Q93073 YZ56_HUMAN - HYPOTHETICAL PROTEIN KIA0256		22278999, 264490, 29331822, 66714117, 66712502, 265006, 265007, 265008, 265009, 264491, 60433438, 265010, 265018, 264760, 264448, 264768, 29148627, 29148628, 265020, 265022, 18108385, 60432113
1866	87268816 (3731, 3732)	Novel Protein sim. GBank gi 5262617 emb CAB45748.1  - (AL080157) hypothetical protein [Homo sapiens]	kinase	18108374, 264769, 18108377, 21906765, 21906766, 35698423, 56182575, 21906769, 29148629, 35698286, 35699917, 265021, 264510, 264511, 264512, 264534, 264535, 60170831, 52644150, 264555, 264691, 60433358, 29331822, 264559, 264595, 264259, 264596, 264692, 264557, 33657023, 29331824, 18108385, 21906754, 33657182, 29331827, 35696032, 33656870, 87168518, 265017, 60431602, 22279000, 264508, 264509, 18108351, 264907, 264682, 264567, 18108372, 264765, 264486
1867	84578159 (3733, 3734)	Novel Protein sim. GBank gi 3859330 (AF078096) - forkhead/winged helix-like transcription factor 7 [Homo sapiens]	UNCLASSIFIED	264094

1868	87357459 (3735, 3736)	Novel Protein sim. GBank gij3881525lembjCAA93884) - (Z70038) cDNA EST EMBL.D32579 comes from this gene; cDNA EST EMBL.D35254 comes from this gene; cDNA EST yk224b3.5 comes from this gene; cDNA EST yk357110.5 comes from this gene [Caenorhabditis elegans]		nuclease	264489, 22278987, 22278999, 29331825, 29331826, 265008, 265009, 33657402, 87168474, 18108351, 21906785, 21906788, 21906769, 265020, 265021, 60170815, 27486264, 264628, 18108374, 264631, 18108385, 87168518, 22279000, 22279002, 284568, 264567
1869	86977282 (3737, 3738)	Novel Protein sim. GBank gij4826772lefnp_004961.1lp(GFA - insulin-like growth factor binding protein, acid labile subunit	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	264508, 264509, 264906, 264908, 264909, 264910, 264591, 264600, 18108351, 264683, 264766, 264769, 35695855, 264634, 264558, 264639, 18108385, 264563, 264486
1870	95349488 (3739, 3740)	Novel Protein sim. GBank gij1869859lembjCAB06722) - (Z86099) very large tegument protein [human herpesvirus 2]		UNCLASSIFIED	29331824, 60424289, 265007, 265008, 21906754, 265017, 265018, 265019, 264288, 264786, 264886, 264888, 21906768, 21906769, 35695917, 80170815, 264692, 18108388, 35695783, 35696423, 65274791, 264638, 264639, 56526486
1871	80234484 (3741, 3742)			UNCLASSIFIED	264509, 264905, 264595, 264768, 264635, 264636, 264563, 264486
1872	80233355 (3743, 3744)	Novel Protein sim. GBank gij2460316 (AF022147) - uterus-ovary specific putative transmembrane protein [Rattus norvegicus]		protease	284510, 264594, 264565
1873	80213890 (3745, 3746)				264509, 264512, 265009, 265011, 18108351, 264687, 264691, 18108370, 18108374, 264635
1874	95351136 (3747, 3748)		Contains protein domain (PF00283) - Bacterial mutT protein	UNCLASSIFIED	264488, 35695917, 264259, 264905, 264907, 264908, 264909, 263978, 264511, 264635, 264638, 264637, 264638, 33637402, 264558, 18108385, 264600, 264804, 264764, 264587, 264788
1875	87330516 (3749, 3750)	Novel Protein sim. GBank gij4589520ldjBAA76782.1) - (AB023155) KIAA0938 protein [Homo sapiens]		UNCLASSIFIED	35696286, 264828, 264582, 264557, 264558
1876	87112860 (3751, 3752)	Novel Protein sim. GBank gij263810lpsl122920 - collagen alpha chain [Riftia pachyplata=tube worms. Peptide, 1027 aa]		UNCLASSIFIED	264259, 29331822, 60432289, 264908, 264909, 264604, 264764, 264288, 264769, 18108376, 264556, 264558, 264559, 18108385
1877	87315208 (3753, 3754)	Novel Protein sim. GBank gij3983355lgbAAC83824.1) - (AF102545) riboflavin binding protein precursor [Scaphiopus couchii]		UNCLASSIFIED	264767, 264886, 264768, 263978, 264693, 264639, 265010, 264563, 264905, 264906, 264907



1878	95351056 (3755, 3756)	Novel Protein sim. GBank gij4510345jgbAAD21434.1] - (AC006921) unknown protein [Arabidopsis thaliana]	Contains protein domain (PF01428) - AN1-like Zinc finger	ubiquitin	264569, 264488, 35696286, 56994075, 264259, 29331822, 29331824, 29331825, 35696052, 29331828, 29146498, 264508, 264509, 264905, 264908, 264907, 264908, 264909, 264510, 264511, 264512, 265009, 264910, 33637402, 264594, 264759, 55812038, 265011, 264602, 264760, 18108351, 264762, 264681, 264682, 264764, 264368, 264288, 264766, 264887, 264768, 264769, 21808766, 35695917, 265021, 60170615, 33657023, 264692, 264693, 33657109, 27486285, 264628, 18108370, 264629, 18108374, 35696423, 264634, 264635, 264555, 264638, 264639, 83373044, 18108385, 56526486, 87188518, 264563, 264564, 264566, 264488, 264567, 264905, 264907, 264908, 265007, 264565, 264566
1879	95310883 (3757, 3758)	Novel Protein sim. GBank gij4928643jgbAAD34082.1]AF15184 - (AF151845) CGI-97 protein [Homo sapiens]		UNCLASSIFIED	
1880	91012978 (3759, 3760)	Novel Protein sim. GBank gij155078jembCAA69283] - (Y08026) Immune associated protein 38 [Mus musculus]		UNCLASSIFIED	264766, 264681, 264682, 83373044
1881	80214949 (3761, 3762)	Novel Protein sim. GBank gij93144jpirjB40505 - hypothetical protein - sud herpesvirus 1 (strain Indiana-Funkhouser or Becker)		UNCLASSIFIED	264509, 264905, 264908, 264909, 264910, 264762, 264687, 33657023, 264632
1882	86582450 (3763, 3764)	Novel Protein sim. GBank gij2384956 (AF022985) - No definition line found [Caenorhabditis elegans]			264908, 21808766, 18108370, 263974, 87188518
1883	94216617 (3765, 3766)	Novel Protein sim. GBank gij1351218jpiP47226]TES2_MOUSE - TESTIN 2 (TES2) [CONTAINS: TESTIN 1 (TES1)]	Contains protein domain (PF00412) - LIM domain containing proteins	homeobox	264908, 264910, 87188559, 21906766, 264636

1884	95310865 (3767, 3768)	Novel Protein sim. GBank gi 4928843 gb AAD34082.1 AF15184 - (AF151845) CGI-87 protein [Homo sapiens]		UNCLASSIFIED	264488, 18108394, 58181686, 22278998, 60432049, 284259, 29331822, 29331824, 29331825, 88714117, 29331827, 35896052, 284508, 284509, 284905, 284906, 284907, 284908, 284909, 58182435, 284510, 284511, 285007, 284512, 285008, 284910, 284592, 284593, 33657402, 5812038, 284758, 5811388, 265010, 265011, 87168559, 284600, 265017, 284604, 265019, 284605, 284760, 5811150, 284761, 284682, 284763, 284683, 284764, 284288, 284369, 284768, 284686, 284768, 284769, 28148784, 35895917, 284680, 284891, 33857023, 284692, 284693, 33857109, 18108370, 284628, 60431528, 284628, 263973, 18108374, 5810764, 5811576, 35896423, 35895855, 284631, 284634, 60431850, 284637, 284638, 58182323, 284639, 18108382, 83373044, 18108385, 60432113, 22279002, 284583, 284584, 284585, 284586, 284486, 284587
1885	87644280 (3789, 3770)	Novel Protein sim. GBank gi 2507155 sp P37370 VRP1_YEAST - VERPROLIN		UNCLASSIFIED	58182575, 284259, 284905, 284909, 285008, 284596, 284766, 265020, 284628, 60431528, 284634, 58526486, 284080, 284583
1886	86674062 (3771, 3772)	Novel Protein sim. GBank gi 2854158 gb AAC02577.1  - (AF045641) No definition line found [Caenorhabditis elegans]			22278998, 22278999, 60432049, 284910, 265018, 284786, 21908768, 29148629, 284690, 284693, 284628, 284555, 284488
1887	94139139 (3773, 3774)	Novel Protein sim. GBank gi 5174421 ref NP_006023.1 pCPNE - copine VI (neuronal)	Contains protein domain (PF00168) - C2 domain	ATPase associated	29331822, 29331824, 29331825, 29331826, 29331827, 284906, 285007, 284681, 284788, 29148627, 284693, 18108364, 35896423, 65274791, 35895855, 284632, 58182323, 284639, 284583
1888	87822804 (3775, 3776)	Novel Protein sim. GBank gi 3319931 emb CAB10841  - (Z98046) dJ1409.2 (Melanoma-Associated Antigen MAGE LIKE) [Homo sapiens]	Contains protein domain (PF01454) - MAGE family		263978
1889	91255783 (3777, 3778)	Novel Protein sim. GBank gi 1083308 pir I56559 - enhancer-trap-locus-1 protein - mouse (fragment)	Contains protein domain (PF00176) - SNF2 and others N-terminal domain	helicase	284906, 5812038, 284758, 265010, 265018, 265019, 18108351, 284288, 21908765, 21908768, 21908769, 265021, 33857023, 33857109, 58182323, 83373044, 18108385, 22279000, 22279002
1890	87626705 (3779, 3780)	Novel Protein sim. GBank gi 4240105 db BAA74876.1  - (AB020660) KIAA0853 protein [Homo sapiens]		UNCLASSIFIED	18108398, 28147620, 284907, 285009, 284600, 265018, 18108351, 284288, 284689, 21908765, 21908768, 21908769, 284691, 284682, 284693, 284628, 18108370, 284636, 284558, 284404

1891	87013895 (3781, 3782)			UNCLASSIFIED	264688, 264768, 264687, 264692, 264693, 29331822, 29331824, 264508, 264905, 264906, 18108370, 264628, 264907, 264908, 264909, 18108379, 265007, 265008, 264910, 264632, 264591, 264639, 264596, 18108384, 265010, 265011, 264601, 264605, 264583, 264369
1892	87642625 (3783, 3784)	Novel Protein sim. GBank gij5689535[dbj][EAA8305.1.1] - (AB029022) KIAA1099 protein [Homo sapiens]	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	silucl	22278995, 264509, 67168559, 18108351, 264448, 264682, 265020, 264693, 18108374, 22279000
1893	86533826 (3785, 3786)			laminin	264589, 65274572, 22278997, 22278999, 264259, 29331822, 29331824, 68714117, 29331826, 264906, 265006, 265008, 265009, 264592, 265018, 264681, 264448, 264683, 18108354, 264368, 264684, 264685, 264766, 264687, 264688, 21906768, 265020, 265022, 60170615, 52644150, 264690, 264691, 264692, 33857023, 264693, 33857109, 264628, 18108374, 35685855, 264630, 264632, 264634, 264557, 264558, 60170394, 18108381, 18108385, 22279000
1894	86599120 (3787, 3788)				264508, 264905, 264906, 264907, 264594, 264684, 264690, 264692, 264630, 264635, 264636, 264639, 264583
1895	87631891 (3789, 3790)	Novel Protein sim. GBank gij5282574[embj][CAB45729.1] - (AL080133) hypothetical protein [Homo sapiens]	Contains protein domain (PF00435) - Spectrin repeat		56182575, 264259, 60432289, 29331826, 264107, 264905, 264908, 264910, 60170831, 264758, 265010, 265018, 264448, 264288, 264768, 33857109, 264628, 55810784, 18108379, 264634, 56182323, 56526486
1896	85673555 (3791, 3792)		Contains protein domain (PF00627) - UBA domain	UNCLASSIFIED	284907, 265008, 264682, 264686, 21906768, 284629, 284631, 284634, 284555
1897	80585589 (3793, 3794)	Novel Protein sim. GBank gij728836[sp]P39193[ALU6_HUMAN . IIII ALU SUBFAMILY SP WARNING ENTRY IIII]		cadherin	264259
1898	87617637 (3795, 3796)	Novel Protein sim. GBank gij127560[sp]P23249[MV10_MOUSE - PROTEIN MOV-10]		helicase	22278996, 22278998, 22278999, 29331824, 29331825, 60432289, 29331827, 35698052, 29331828, 265008, 265019, 264681, 264682, 264448, 264368, 52644229, 21906765, 21906766, 21906768, 21906769, 60170615, 55810784, 22279000
1899	86673097 (3797, 3798)	Novel Protein sim. GBank gij2909819 (AF031548) - erythrocyte membrane glycoprotein Rh50 [Homo sapiens]	Contains protein domain (PF00909) - Ammonium Transporter Family	glycoprotein	264259, 264508, 264908, 60432228, 264768, 21906765, 21906769
1900	87641858 (3799, 3800)	Novel Protein sim. GBank gij4102881 (AF017250) - vitellogenin precursor [Oreochromis aureus]		UNCLASSIFIED	264683

1901	95196647 (3801, 3802)	Novel Protein sim. GBank gij58599spjP38378[S61A_RAT PROTEIN TRANSPORT PROTEIN SEC61 ALPHA SUBUNIT	Contains protein domain (PF00242) - DNA polymerase (viral) N-terminal domain	transport	264488, 52644507, 52645156, 18108398, 52646385, 52648642, 18108397, 56182575, 22278894, 22278895, 56994075, 35688286, 22278897, 22278898, 264490, 60432049, 264259, 29331822, 52645080, 29331824, 29331825, 66714117, 28331826, 60432289, 29331827, 29331828, 35696052, 31858970, 29146498, 264906, 264907, 29331830, 264908, 52644045, 264909, 264112, 265006, 264512, 265008, 264910, 265009, 60170831, 60432229, 60433356, 33657402, 60433438, 55812038, 264756, 33109954, 21908754, 33657084, 52644296, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 18108351, 264448, 264288, 264686, 21906769, 55811957, 35695917, 265020, 52644229, 21906765, 21906766, 21906767, 265021, 52644190, 18108362, 33657023, 264683, 263887, 33657109, 33657182, 27488264, 33657349, 35885763, 18108370, 18108376, 55811576, 35696423, 35695855, 60431850, 264636, 263981, 52644332, 60170394, 83373044, 18108385, 87168518, 60432113, 264564, 264107, 263976
1902	80202013 (3803, 3804)	Novel Protein sim. GBank gij44266[3]gbjAAD20451[- (AF098799) SLM-1 [Mus musculus]		dna_mta_bind	
1903	87778554 (3805, 3806)	Novel Protein sim. GBank gij3747107 (AF095741) - unknown [Rattus norvegicus]		UNCLASSIFIED	264259, 29331825, 29331827, 264508, 264907, 265008, 60170831, 60433356, 60433438, 264759, 21906754, 264448, 264288, 265021, 265022, 33657023, 264693, 55811576, 264555, 264556, 22279000
1904	80434213 (3807, 3808)	Novel Protein sim. GBank gij1332911spjP47147YJ80_YEAST - HYPOTHETICAL 80.2 KD PROTEIN IN CPA2-NNF1 INTERGENIC REGION		sinud	284509, 264805, 264806, 264507, 264908, 285007, 264810, 264686, 264768, 264687, 264789, 264693, 264628, 18108374, 264634, 264638, 264637, 264583
1905	85351140 (3809, 3810)	Novel Protein sim. GBank gij3043714[dbj]BAA25521[- (AB011167) KIAA0595 protein [Homo sapiens]	Contains protein domain (PF00293) - Bacterial mutT protein		264488, 264768, 264769, 264689, 29146629, 35695917, 35696286, 264259, 264692, 18108362, 33657023, 29331824, 33657109, 29146499, 264508, 264509, 264905, 264906, 264907, 66712502, 264908, 264909, 35886423, 35685855, 264510, 264511, 264512, 264910, 264634, 264635, 264637, 264638, 33657402, 264756, 85655542, 264602, 264760, 264761, 264482, 264563, 264762, 264483, 264764, 264566, 264288, 264765
1906	12763822 (3811, 3812)			UNCLASSIFIED	264637

1907	95351144 (3813, 3814)	Novel Protein sim. GBank gi 49295895 gb AAD34053.1 AF15181 - (AF151816) CGI-58 protein [Homo sapiens]	Contains protein domain (PF00561) - hydrolase alpha/beta hydrolase fold	UNCLASSIFIED	55274572, 22278998, 35698286, 22278998, 22278998, 264259, 80432049, 29331822, 29331825, 60424259, 29331828, 35696052, 284593, 60433356, 21906754, 55811388, 85658542, 87188559, 265018, 264681, 264682, 264684, 264288, 21906765, 21906766, 21906768, 265020, 265022, 264690, 52644150, 264692, 33857023, 264693, 33857109, 35695855, 264636, 264638, 60432113
1908	95313641 (3815, 3816)	Novel Protein sim. GBank gi 3986770 (AF109906) - NG22 [Mus musculus]		UNCLASSIFIED	264488, 55274572, 56182575, 56181686, 22278995, 22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 60432289, 29331826, 29331827, 29331828, 29146499, 264905, 264906, 264907, 264908, 86712502, 264909, 56182435, 264510, 264511, 265008, 264910, 60170831, 264592, 60433356, 33657402, 264594, 60433438, 264596, 55812038, 33109954, 52646317, 265011, 265017, 264604, 265018, 265019, 264605, 55811150, 264681, 264448, 264288, 264686, 264688, 264769, 21906765, 21906766, 21906767, 21906768, 21906769, 55811857, 29146628, 35695917, 265020, 265022, 264691, 264692, 18108364, 85274620, 33657109, 33657348, 35695763, 18108374, 263978, 55810764, 55811576, 35696423, 65274791, 284631, 284632, 284556, 264557, 60170394, 56182323, 83373044, 18108385, 60432113, 22279000, 22279002, 264566, 264486
1909	85514505 (3817, 3818)	Novel Protein sim. GBank gi 2224553 db BAA20813  - (AB002354) KIA0356 [Homo sapiens]		UNCLASSIFIED	284258, 264508, 264905, 264906, 264907, 264908, 264511, 264910, 264593, 264758, 264764, 264766, 18108370, 264634, 264637, 284488
1910	94216621 (3819, 3820)	Novel Protein sim. GBank gi 1351218 sp P47226 TES2_MOUSE - TESTIN 2 (TES2) [CONTAINS: TESTIN 1 (TES1)]	Contains protein domain (PF00412) - homeobox LM domain containing proteins	homeobox	35695285, 22278996, 22278999, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 285007, 284512, 264910, 264758, 285011, 264601, 264602, 264604, 264605, 264761, 264764, 264288, 264768, 284768, 264687, 264769, 35695917, 265021, 52644150, 264692, 264628, 18108370, 284629, 18108372, 18108374, 35695855, 264631, 264634, 264635, 264636, 264637, 264638, 18108385, 284585, 284586, 284488

1911	81725345 (3821, 3822)	Novel Protein sim. GBank gjl4809339igp/AD30184.1(AC006530) hypothetical protein [Homo sapiens]	Contains protein domain (PF01119) - DNA mismatch repair protein	nuclease	18108394, 56182575, 56182181, 29331826, 29331827, 33656970, 264906, 265007, 264591, 55812038, 87166559, 264448, 264369, 21906765, 21906768, 265022, 264691, 264693, 18108395, 55811576, 264556, 18108395, 18108398
1912	85413519 (3823, 3824)	Novel Protein sim. GBank gjl5699439dbj/BAA83003.1] - (AB028974) KIAA1051 protein [Homo sapiens]	Contains protein domain (PF00098) - Zinc finger, CCHC class	UNCLASSIFIED	18108397, 56182575, 56181686, 22278994, 22278995, 56994075, 22278996, 22278999, 264259, 29331822, 29331824, 56182181, 29331825, 66714117, 35686052, 264903, 264906, 264907, 264908, 52644045, 56182435, 265007, 265008, 264910, 265009, 264591, 264596, 65274444, 55811386, 87168474, 265011, 87168559, 265018, 265019, 264760, 18108351, 264681, 264369, 264684, 264288, 264686, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265020, 265021, 265022, 60170615, 264692, 33657023, 264693, 18108376, 55811576, 35686423, 85274791, 264637, 56182323, 83373044, 56528486, 22279002, 264563, 264566
1913	85305546 (3825, 3826)	Novel Protein sim. GBank gjl5032245jrefj/NP_005665.1p2NF2 - zinc finger protein (C2H2) homologous to mouse MOK-2			56182575, 22278994, 22278995, 56994075, 22278996, 22278998, 22278999, 29331826, 29331827, 265006, 55812038, 265010, 265017, 265018, 265018, 264681, 18108351, 264683, 264764, 264369, 264288, 264685, 264686, 264769, 21906765, 21906766, 21906768, 21906769, 55811957, 265020, 265022, 264691, 55811576, 264634, 264635, 264638, 56182323, 83373044, 18108385
1914	83423982 (3827, 3828)	Novel Protein sim. GBank gjl4589604dbj/BAA76824.1] - (AB023187) KIAA0980 protein [Homo sapiens]	Contains protein domain (PF00036) - EF hand	struct	56182575, 29331824, 35696052, 264906, 264908, 264828, 264909, 264592, 264758, 87168559, 18108351, 18108354, 264684, 264686, 33657023, 264693, 264628, 264631, 264632, 264634, 264635, 264639
1915	95340459 (3829, 3830)	Novel Protein sim. GBank gjl5689415dbj/BAA82991.1] - (AB028962) KIAA1039 protein [Homo sapiens]		UNCLASSIFIED	264259, 29331824, 29331826, 29331827, 264508, 264909, 265009, 265017, 265019, 264768, 264769, 264689, 264628, 264635, 264637, 264639, 83373044, 264585
1916	79640761 (3831, 3832)				264693, 264639

1917	87821880 (3833, 3834)	Novel Protein sim. GBank gij568939.1[dbj BA82979.1] - (AB028950) KIAA1027 protein [Homo sapiens]		struct	264769, 264689, 21906765, 21906766, 22278996, 264259, 264691, 264693, 29331824, 29331825, 29331826, 29331828, 284905, 284906, 264628, 264807, 264808, 284908, 264510, 264830, 264910, 264634, 264635, 264636, 264637, 264638, 263981, 264639, 264758, 18108385, 21906754, 265011, 264604, 264583, 18108351, 264762, 264763, 264566, 264764, 264766
1918	95302795 (3835, 3836)	Novel Protein sim. GBank gij5281517[gb AA041524.1]AF15483 - (AF15483) PV-1 [Rattus norvegicus]		struct	264488, 18108392, 18108357, 21906765, 21906767, 21906768, 56182575, 21908769, 22278994, 35696286, 35695917, 22278996, 22278997, 265021, 265022, 264534, 264690, 264691, 264692, 33657023, 264693, 29331824, 29331825, 33657109, 29331826, 52645129, 35696052, 29331828, 27486262, 27486264, 35695763, 264508, 264905, 264508, 264908, 264628, 264907, 18108370, 264908, 264829, 264909, 18108372, 18108374, 263978, 35696423, 35695855, 264510, 264511, 265006, 265007, 264512, 265008, 264631, 265009, 264910, 264634, 264635, 264555, 264636, 264556, 264637, 264557, 264593, 264638, 264594, 60170394, 264595, 264559, 264596, 83373044, 264759, 52646317, 18108385, 52644296, 56528486, 87168518, 265010, 265011, 87168559, 264600, 264601, 264602, 265017, 264603, 264604, 265018, 264605, 264760, 264761, 264482, 264584, 18108351, 264762, 264692, 264565, 264448, 264764, 264566, 264486, 264567, 264369, 264288, 264766, 264487, 264685
1919	94143847 (3837, 3838)	Novel Protein sim. GBank gij3878594[emb CA01237] - (Z77667) cDNA EST EMBL:CO8125 comes from this gene; cDNA EST EMBL:CO9753 comes from this gene [Caenorhabditis elegans]		oxidase	22278997, 29331822, 265007, 60170831, 60432229, 60433438, 264448, 264682, 264288, 55811957, 33657023, 33657109, 65274791, 56182323, 22279002
1920	91228953 (3839, 3840)	Novel Protein sim. GBank gij1809231 (AC000115) - coded for by human cDNAs R78043 (NID:g850725), R65857 (NID:g838495) and H12868 (NID:g877688) [Homo sapiens]		UNCLASSIFIED	264510, 264511, 264512, 264568
1821	79555226 (3841, 3842)	Novel Protein sim. GBank gij4580997[gb AAD24571.1]AF12108 - (AF12108) cAMP inducible 2 protein [Mus musculus]		UNCLASSIFIED	264693

1922	87641863 (3843, 3844)	Novel Protein sim. GBank gi 138595 sp P02845 VIT2_CHICK - VITELLOGENIN II PRECURSOR (MAJOR VITELLOGENIN) (CONTAINS: LIPOVITELLIN I (LVI); PHOSVITIN (PV); LIPOVITELLIN II (LVII); YGP40)		UNCLASSIFIED	264586, 264588, 264490, 18108370, 264909, 18108374, 265008, 264557, 264564, 18108351
1923	94323589 (3845, 3846)	Novel Protein sim. GBank gi 119110 sp P03211 EBN1_EBV - EBNA-1 NUCLEAR PROTEIN		UNCLASSIFIED	264488, 264489, 22278985, 264094, 264259, 35696052, 264509, 264905, 264906, 264807, 264908, 66712502, 264909, 264511, 264512, 265007, 264910, 265009, 264591, 264592, 264593, 264594, 264595, 264758, 264598, 264759, 265011, 265017, 265018, 265019, 55911150, 264681, 264762, 264448, 264764, 264288, 264369, 264766, 264767, 264686, 264687, 264768, 264769, 56181552, 264689, 21906766, 264691, 33657023, 264693, 65274620, 33657109, 18108370, 264628, 264629, 35698423, 264830, 264631, 264632, 264634, 264555, 264636, 264637, 264638, 264639, 264558, 83373044, 87168518, 264563, 264564, 264565, 264566, 264567, 26331826, 264806, 264908, 264595, 265017, 265018, 265019, 265021, 264691, 264693, 264637, 18108385, 264565
1924	87338925 (3847, 3848)	Novel Protein sim. GBank gi 3877655 emb CA96657  - (Z72511) possible zinc finger protein; cDNA EST EMBL:M89115 comes from this gene; cDNA EST EMBL:D71533 comes from this gene; cDNA EST EMBL:D72314 comes from this gene; cDNA EST EMBL:D75164 comes from this gene; cDNA EST EMBL:C1....	Contains protein domain (PF00087) - Zinc finger, C3HC4 type (RING finger)		
1925	87628338 (3849, 3850)	Novel Protein sim. GBank gi 4981903 gb AAD36415.1 AE001788 - (AE001788) ribosomal protein S15 (Thermotoga maritima)	Contains protein domain (PF00312) - Ribosomal protein S15	ribosomalprot	22278995, 22278996, 22278997, 264259, 29331824, 66714117, 26146499, 264909, 52044045, 265008, 265009, 264756, 265011, 265017, 264605, 264446, 264288, 264692, 33657108, 18108374, 60170394
1926	88094739 (3851, 3852)	Novel Protein sim. GBank gi 2246532  (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma- associated herpesvirus]		struct	264905, 264906, 264907, 264910, 264593, 265018, 264760, 264764, 264288, 264692, 264693, 263978, 264631, 264634, 264637, 264563
1927	85654857 (3853, 3854)	Novel Protein sim. GBank gi 3043632 dbj BA025480  - (AB011126) KIA0554 protein [Homo sapiens]			264689, 264631
1928	87798054 (3855, 3856)	Novel Protein sim. GBank gi 1655761 dbj BA013377  - (D87433) KIA00249 [Homo sapiens]	Contains protein domain (PF00193) - Inf Extracellular link domain	Inf	264489, 264259, 265017, 265021, 264692
1929	86997235 (3857, 3858)	Novel Protein sim. GBank gi 5001993 gb AAD37247.1 AF13432 - (AF13432) chimeric AFGP/hypsinogen-like serine protease precursor [Dissostichus mawsoni]		UNCLASSIFIED	264508, 264591, 33657402, 265017, 264768, 264632, 264556, 264639



1930	87889128 (3859, 3860)	Novel Protein sim. GBank gil1709230 sp P52963 NBL4_MOUSE - NBL4 PROTEIN			phosphatase	35696286, 29331828, 264905, 264907, 264908, 264909, 264511, 264910, 264758, 264601, 265017, 265019, 264605, 264760, 264764, 264766, 264686, 264769, 265022, 35696423, 264638, 60432113
1931	87767270 (3861, 3862)	Novel Protein sim. GBank gij404634 (U01840) - serine/threonine kinase [Mus musculus]		Contains protein domain (PF00089) - Eukaryotic protein kinase domain	kinase	284906, 264908, 60432228, 264758, 264764, 264288, 265020, 284692, 264634, 284637, 284684, 284691, 284635
1932	15030972 (3863, 3864)				UNCLASSIFIED	284595
1933	11613668 (3865, 3866)				struct	56182575, 56182435, 264510, 264757, 264758, 55812038, 55811386, 265018, 55811150, 21908765, 264681, 264831, 264635, 264637
1934	84426360 (3867, 3868)	Novel Protein sim. GBank gil4115748 dbj BAA36484  - (A022023) nonmuscle myosin heavy chain B [Bos taurus]			UNCLASSIFIED	284686, 285011, 284511, 264905, 18108351, 264564, 264681, 264258, 18108370, 284586, 264764, 264369, 284595
1935	87752511 (3869, 3870)					60432289, 265007, 265010, 285011, 285018, 33657109, 18108374
1936	95414338 (3871, 3872)	Novel Protein sim. GBank gil4827040 ref NP_005110.1 pTRAP - thyroid hormone receptor-associated protein, 150 kDa subunit		Contains protein domain (PF00023) - Ank repeat	kinase	35558542, 21908767, 35695917, 60170615, 284683, 33657109
1937	94847141 (3873, 3874)	Novel Protein sim. GBank gil543187 pit J53771 - ankyrin, erythrocyte - mouse		Contains protein domain (PF00888) - Cullin family	collagen	264488, 29146408, 264905, 264559
1938	87403277 (3875, 3876)	Novel Protein sim. GBank gil4544431 gb AAD22340.1 AC006955 - (AC006955) hypothetical protein [Arabidopsis thaliana]				
1939	81004978 (3877, 3878)	Novel Protein sim. GBank gil500858 dbj BAA03210  - (D14188) 50kDa lectin [Bombyx mori]			UNCLASSIFIED	65274572, 22278999, 60432289, 56182435, 60433356, 265017, 21908765, 21908768, 21908768, 55811957, 27486264, 35696423, 60432113, 264584
1940	87348810 (3879, 3880)	Novel Protein sim. GBank gil1945300 emb CAA73132  - (Y12529) hypothetical protein [Silene latifolia]		Contains protein domain (PF00560) - Leucine Rich Repeat	struct	264488, 29331822, 264448, 264683, 264288, 265020, 33657023, 264631
1941	94147177 (3881, 3882)	Novel Protein sim. GBank gil4206386 (AF060570) - rig-1 protein [Mus musculus]			UNCLASSIFIED	56994075, 22278998, 264259, 29331824, 29331827, 264905, 265008, 33657084, 265017, 265018, 264288, 264687, 21908765, 21908766, 21908767, 265020, 52644150, 27488284, 83373044, 18108387, 60432113, 22279002, 284585
1942	87841870 (3883, 3884)	Novel Protein sim. GBank gil4927204 gb AAD33049.1 AF133911 - (AF133911) ARL-6 interacting protein-4 [Mus musculus]			UNCLASSIFIED	264488, 18108388, 29331825, 27486261, 264509, 18108370, 18108374, 264482
1943	94325298 (3885, 3886)	Novel Protein sim. GBank gil3122952 sp O15736 TIPD_DICDI - TIPD PROTEIN		Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	22278998, 29331822, 29331827, 35696052, 264511, 265009, 264592, 60432229, 265017, 265018, 265019, 264684, 284692, 33657109, 65274791, 264636

1844	94232958 (3887, 3888)	Novel Protein sim. GBank gll1798570[dbj]BAA134321 - (D87671) TIP120 [Rattus norvegicus]	UNCLASSIFIED	65274572, 22278994, 22278995, 22278996, 22278998, 22278999, 264259, 52645080, 29331824, 29331826, 29331827, 66712502, 56182435, 264512, 264910, 60170831, 60432229, 60433356, 33657402, 60433438, 264596, 33109954, 21908754, 87168474, 87168559, 265017, 265018, 265019, 18108351, 264369, 264686, 264766, 21908765, 21908766, 21908767, 21908769, 35695917, 265020, 265021, 60170615, 264692, 33657023, 18108370, 18108374, 35696423, 35695855, 264634, 60170394, 264639, 83373044, 18108385, 56526486, 87168518, 60432113, 22279000, 264563
1945	87641872 (3889, 3890)	Novel Protein sim. GBank gll4927204[gb]AAD33049.1[AF13391] ARL-6 interacting protein-4 [Mus musculus]	UNCLASSIFIED	264488, 22278996, 264510, 264511, 18108351, 264683, 264486, 264507
1946	87443990 (3891, 3892)	Novel Protein sim. GBank gll2498104[sp]Q27989[AD50_BOVIN - ADRENAL MEDULLA 50 KO PROTEIN]	UNCLASSIFIED	60432289, 29331827, 35696052, 265007, 265008, 60433356, 60433438, 264369, 56181562, 21906767, 52644150, 264693, 27486264, 264637, 87168518, 264563
1947	86438862 (3893, 3894)	Novel Protein sim. GBank gll3914801[sp]Q54888[RP2_RAT - DNA-DIRECTED RNA POLYMERASE I 135 KD POLYPEPTIDE (RNA POLYMERASE I SUBUNIT 2) (RPA135) (RNA POLYMERASE I 127 KD SUBUNIT)]	napolymerase	22278998, 264905, 264906, 264908, 264909, 264512, 284758, 264762, 264682, 264683, 264764, 264288, 264788, 264888, 21906788, 264693, 18108374, 35695855, 264635, 264637, 264638, 18108385, 22279002
1948	95199174 (3895, 3896)	Novel Protein sim. GBank gll5420387[emb]CAB46679.11 - (AJ243459) proteophosphoglycan [Leishmania major]	struc1	264909, 60170831, 264591, 264594, 235010, 265011, 264764, 264369, 264689, 264631, 264638
1949	7640129 (3897, 3898)			264369
1950	87788531 (3899, 3900)	Novel Protein sim. GBank gll3876766[emb]CA93466.11 - (Z69637) predicted using Genefinder; Similarity to E.coli hypothetical protein YCAC (SW:YCAC_ECOLI) [Caenorhabditis elegans]	UNCLASSIFIED	264488, 264788, 264688, 264689, 264259, 29331822, 33657109, 29331828, 264103, 264509, 18108370, 35695855, 264510, 265008, 265009, 33657402, 18108385, 265018, 264563, 55811150, 18108351, 264369, 264288, 18108354
1951	86988253 (3901, 3902)	Novel Protein sim. GBank gll2626753[dbj]BAA234241 - (AB008782) sulfate transporter [Arabidopsis thaliana]	transport	56182575, 22278997, 52645080, 29331824, 29331825, 29331827, 55812038, 52646317, 265018, 265019, 264369, 21906785, 21906767, 55311857, 285020, 285021, 33657023, 264693, 35695763, 58182323, 22279002
1952	87069775 (3903, 3904)	Novel Protein sim. GBank gll4929833[gb]AAD34077.1[AF15184] CGI-82 protein [Homo sapiens]	reductase	264591, 264593, 264594, 264595, 264555, 264556, 264557, 264558, 264565

1953	20470371 (3905, 3906)	Novel Protein sim. GBank gij1168715ipjP31721[C1QB_RAT - COMPLEMENT C1Q SUBCOMPONENT, B CHAIN PRECURSOR (AB020698) KIAA0891 protein (Homo sapiens)]	Contains protein domain (PF00386) - C1q domain	complement	264259, 264558
1954	91226025 (3907, 3908)	Novel Protein sim. GBank gij4240271[dbjBAA74914.7] - (AB020698) KIAA0891 protein (Homo sapiens)]	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	264488, 263994, 18108394, 35696286, 22278998, 29331822, 66714117, 29331828, 29331827, 264508, 264509, 264905, 264906, 284807, 284908, 264809, 264510, 265006, 265007, 265008, 284810, 265009, 60170831, 60432229, 265011, 265017, 264603, 264604, 265018, 18108351, 264685, 264766, 264769, 35695917, 265020, 265021, 264691, 33657023, 264692, 33657109, 264628, 18108374, 35696423, 55811576, 35695855, 264630, 264635, 264636, 264555, 264556, 264638, 264557, 264639, 264558, 264559, 83373044, 18108385, 264563, 264564, 264568, 264488, 264567
1955	95308310 (3909, 3910)			UNCLASSIFIED	56182575, 56181688, 22278994, 22278999, 264259, 29331822, 56182181, 29331824, 29331825, 29331826, 29331827, 35696052, 264508, 29331830, 265008, 265009, 264581, 55812038, 87169474, 265017, 265018, 265018, 264448, 284786, 21905785, 21906768, 21906787, 55811957, 265020, 265021, 52645129, 33657109, 27489284, 33657349, 35695763, 60431528, 18108374, 55811576, 35695855, 264635, 60431850, 264639, 83373044
1956	95092121 (3911, 3912)	Novel Protein sim. GBank gij1665821[dbjBAA13407] - (D87469) Similar to D.melanogaster cadherin-related tumor suppressor (Homo sapiens)]	Contains protein domain (PF00028) - Cadherin domain	cadherin	264488, 56182575, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 264508, 264906, 264908, 264909, 264511, 265007, 264910, 264591, 55812038, 265010, 265018, 18108351, 264768, 56181562, 264689, 21906768, 21906789, 265022, 264691, 264628, 18108374, 55810764, 55811576, 35695855, 264631, 264632, 264635, 264637, 264638, 60170394, 56182323, 83373044, 18108385, 22278000, 22278002, 264563, 264564

1957	94326510 (3913, 3914)	Novel Protein sim. GBank gi4589674[dbj][BAA78856.1] - (AB023229) KIAA1012 protein [Homo sapiens]		UNCLASSIFIED	52646842, 58182575, 22278997, 22278998, 22278999, 29331824, 68714117, 29331827, 29146498, 264593, 33657402, 33108954, 87168474, 265018, 264448, 264368, 264288, 264766, 21906766, 21908767, 21806768, 21906769, 265020, 265021, 264992, 65274620, 27486264, 33657349, 27486265, 35859555, 22279002, 264482
1958	95313902 (3915, 3916)	Novel Protein sim. GBank gi4240227[dbj][BAA74892.1] - (AB020676) KIAA0869 protein [Homo sapiens]		UNCLASSIFIED	22278999, 264092, 264094, 264259, 80432049, 29331824, 56182181, 68714117, 264107, 264109, 264909, 264511, 60170831, 60432229, 21906754, 265010, 21906769, 35859517, 265022, 65274620, 263967, 263976, 35868423, 264831, 264632, 264634, 264635, 18108385, 22279000, 22279002, 264593, 265019
1959	85701470 (3917, 3918)	Novel Protein sim. GBank gi2281983[emb][CAB10860] - (Z96056) hypothetical protein [Schizosaccharomyces pombe]		ubiquitin	
1960	80308608 (3919, 3920)	Novel Protein sim. GBank gi2274851[dbj][BAA21515] - (D64159) 3'-7 gene product [Homo sapiens]		struct	264905, 264906, 264907, 264908, 264909, 265006, 265007, 264910, 264595, 265017, 264604, 265018, 18108351, 264764, 264369, 264766, 264768, 21906765, 18108368, 264629, 18108379, 264635, 264636, 264637, 264638, 264486
1961	16292607 (3921, 3922)				264635
1962	91008385 (3923, 3924)				
1963	90936017 (3925, 3926)	Novel Protein sim. GBank gi3721653[dbj][BAA33581] - (AB012933) acyl-CoA synthetase 5 [Rattus norvegicus]		UNCLASSIFIED	85274572, 264592, 264593, 265019, 264691, 85274572, 18108398, 35868286, 29331825, 60432289, 29331827, 264828, 265006, 265009, 60433356, 60433438, 21906754, 265020, 265021, 33857023, 33657109, 27486265, 35859555, 264555
1964	94317605 (3927, 3928)	Novel Protein sim. GBank gi5262638[emb][CAB45757.1] - (AL080169) hypothetical protein [Homo sapiens]		cadherin	264488, 264092, 264259, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265007, 265009, 264910, 264592, 264593, 264594, 264595, 264758, 264600, 264603, 264604, 264605, 264760, 264762, 264448, 264764, 264288, 264685, 264768, 264768, 264769, 21908766, 264691, 264692, 264693, 18108370, 264628, 264629, 18108374, 264830, 264631, 264634, 264637, 264638, 18108382, 83373044, 18108385, 264483, 264584, 264565, 264566, 264486, 284567
1965	94317445 (3929, 3930)	Novel Protein sim. GBank gi4107017[dbj][BAA36294] - (AB001773) PEM-6 [Ciona savignyi]	Contains protein domain (PF01428) - AN1-like Zinc finger	ubiquitin	264488, 264510, 264760, 264768, 264486

1866	94192058 (3831, 3832)	Novel Protein sim. GBank gi 4929707 gb AAD34114.1 AF15187 - (AF15187) CGI-119 protein [Homo sapiens]	Contains protein domain (PF01027) - Uncharacterized protein family	glycoprotein	22278999, 264092, 264259, 29331826, 29331828, 29146498, 264895, 265011, 264448, 18108354, 264288, 264684, 264766, 264885, 264886, 265022, 264691, 264692, 18108370, 18108377, 264555, 18108381, 18108385, 264486, 264587
1867	87396123 (3833, 3934)	Novel Protein sim. GBank gi 2957270 (AF044576) - phospholipase C PLC210 [Caenorhabditis elegans]	Contains protein domain (PF00388) - Phospholipase C, X domain	esterase	29331824, 265010, 265017, 264288, 21906764, 263981, 56526486
1868	88095641 (3835, 3936)	Novel Protein sim. GBank gi 2584953 (AF030001) - unknown [Mus musculus]	Contains protein domain (PF00008) - EGF-like domain	oncogene	35698286, 264805, 264509, 264908, 264807, 264808, 264809, 264511, 264512, 285008, 264910, 265009, 264594, 264757, 264758, 264604, 264605, 264760, 264762, 264682, 264764, 264685, 264768, 264767, 264689, 264691, 264693, 264628, 264629, 35696423, 35695855, 264631, 264632, 264634, 264635, 264636, 264637, 18108380, 264584, 264585, 264588, 264587
1869	84328529 (3837, 3938)	Novel Protein sim. GBank gi 2911274 (U20329) - spidroin 1 [Nephila clavipes]		UNCLASSIFIED	22278995, 22278996, 35696052, 264906, 264908, 18108351, 264482
1870	80596049 (3839, 3940)	Novel Protein sim. GBank gi 4050087 (AF109907) - S164 [Homo sapiens]		UNCLASSIFIED	264908, 264288, 264766, 264636
1871	84843914 (3941, 3942)	Novel Protein sim. GBank gi 134208 pp P09593 SANT_PLAFV - SANT-GEN PROTEIN PRECURSOR		collagen	264488, 264489, 22278998, 264259, 60432049, 66714117, 29331826, 60432289, 29331827, 35696052, 264508, 264905, 294509, 264908, 264907, 264908, 264909, 264510, 264511, 264512, 264910, 264591, 264592, 60432229, 60433356, 264595, 264596, 264600, 264604, 264605, 264760, 18108351, 264448, 264764, 264288, 264766, 264768, 264769, 21906765, 33637023, 264692, 18108370, 264629, 35696423, 65274791, 35695855, 264632, 264635, 264555, 264638, 264637, 264638, 264639, 18108385, 60432113, 22279000, 264583, 264584, 264555, 264556, 264486
1872	87645444 (3843, 3944)	Novel Protein sim. GBank gi 4519623 db BAA75671.1  - (AB017616) homologous to the yeast YGR163 gene [Mus musculus]	Contains protein domain (PF01462) - Leucine rich repeat N-terminal domain		22278999, 264259, 29331822, 56182181, 60432289, 29331827, 52644045, 264909, 265006, 264511, 265008, 52644296, 265018, 265019, 264761, 264889, 21906768, 21906769, 264691, 264693, 33637109, 33637182, 264558, 52644332, 264558, 60432113
1873	86395533 (3945, 3946)			UNCLASSIFIED	29331826, 264692, 35696423, 264631, 264555, 264556, 264557, 264558, 264559
1874	80396629 (3947, 3948)	Novel Protein sim. GBank gi 3309543 (AF036382) - ML1 [Fugu rubripes]		UNCLASSIFIED	264682, 264764, 264563

1875	94316479 (3949, 3950)				UNCLASSIFIED	264488, 66714117, 29331826, 29331828, 56182435, 265006, 264757, 55812038, 265010, 265017, 264369, 55811957, 65274781, 35695855, 56182323, 60432113
1876	95355914 (3951, 3952)				UNCLASSIFIED	264259, 35696052, 265018, 265020, 265021, 33657109, 56526486
1877	94855664 (3953, 3954)				homeobox	264908, 264596, 265021, 264566
1878	87447645 (3955, 3956)				transcript factor	60170831, 264566
1879	87627709 (3957, 3958)				UNCLASSIFIED	29331826, 29146498, 264905, 264907, 265007, 265009, 265010, 265018, 264686, 18108359, 21906768, 35695917, 265020, 60170815, 264693, 18108368, 18108370, 264631, 264635, 264556, 264558, 18108384, 22279000, 264565
1880	86577059 (3959, 3960)				ubiquitin	264489
1881	87606974 (3961, 3962)				UNCLASSIFIED	22278995, 22278997, 22278999, 264259, 29331825, 29331828, 29146498, 29146499, 264107, 264908, 264910, 264595, 21906734, 265010, 265017, 265018, 265019, 264448, 264288, 21906767, 33657023, 27486264, 18108370, 18108374, 18108376, 264630, 264631, 264635, 18108385, 87168518, 22279000, 264482, 264564
1882	80895367 (3963, 3964)					65274572, 29331825, 35696052, 33656970, 264909, 265008, 55811386, 264780, 284686, 264691, 27486264
1883	95088668 (3965, 3966)				transcript factor	22278995, 35696286, 264259, 29331822, 20281099, 29331824, 60432289, 29331827, 264509, 264905, 264906, 264907, 66712502, 264908, 52644045, 264909, 264510, 264512, 264910, 265009, 264591, 264592, 60433356, 60433438, 264758, 265010, 264600, 264603, 264604, 264760, 264762, 264763, 264764, 264766, 264887, 264768, 264769, 21908765, 55811957, 35695817, 264690, 264682, 264693, 284628, 264629, 263978, 18108379, 35696423, 35695855, 20281071, 264632, 264634, 264635, 264555, 264636, 264637, 264638, 264639, 63373044, 18108385, 22279000, 22279002

1884	85760889 (3967, 3968)	Novel Protein sim. GBank gij2886695(embjCAA17174.1) - (AL021897) fadD14 [Mycobacterium tuberculosis]		synthase	264688, 21906768, 55811957, 56994075, 265020, 265021, 22278999, 265022, 264239, 29331822, 33657182, 29146499, 264628, 18108370, 264908, 264628, 55811576, 35695855, 265008, 265007, 264591, 21906754, 33657084, 265010, 265017, 265019, 264288
1885	85638887 (3969, 3970)	Novel Protein sim. GBank gij5712131(gb AAD47379.1 AF12049) DEM1 protein [Homo sapiens]		glycoprotein	264760, 264288, 263978, 55811576, 264637, 56182323, 18108385, 264594
1886	80200507 (3971, 3972)	Novel Protein sim. GBank gij4868443(gb AAD31319.1 AF14457) Mx-interacting protein kinase PKM [Mesocricetus auratus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	UNCLASSIFIED	264488, 264629
1887	87011117 (3973, 3974)	Novel Protein sim. GBank gij4868443(gb AAD31319.1 AF14457) Mx-interacting protein kinase PKM [Mesocricetus auratus]		UNCLASSIFIED	22278999, 29331830, 265007, 265018, 21906768, 33657023, 264692, 264693, 18108377, 264635, 60170394, 22279002
1888	94122108 (3975, 3976)			UNCLASSIFIED	264905, 264906, 264907, 264908, 264909, 264910, 264591, 264593, 264758, 264764, 264686, 264768, 265021, 264692, 264628, 264629, 35695855, 264630, 264635, 264636, 264637, 264638, 264639, 264483
1889	91225225 (3977, 3978)	Novel Protein sim. GBank gij2801701 (AF042379) - spindle pole body protein spc97 homolog GCP2 [Homo sapiens]		tubulin	60432049, 60432289, 52644045, 56182435, 264112, 265007, 33657402, 52644229, 21906768, 21906768, 21906768, 55811857, 33657023, 263967, 33657109, 18108370, 22279000, 22279002
1890	85698888 (3978, 3980)	Novel Protein sim. GBank gij5701727(dbj BAA83074.1) - (AB024728) alpha-1,3-D-mannoside beta-1,4-N-acetylglucosaminyltransferase IV-homologue [Homo sapiens]			264508, 264757, 264764, 18108381
1891	95353114 (3981, 3982)	Novel Protein sim. GBank gij4240287(dbj BAA74922.1) - (AB020708) KIAA0899 protein [Homo sapiens]	Contains protein domain (PF01602) - Adaptin N terminal region	glycoprotein	18108394, 56182575, 22278994, 35688288, 56994075, 22278997, 22278999, 29331822, 29331824, 29331825, 60432289, 29331828, 264508, 264908, 264907, 264908, 56182435, 264510, 265007, 21906754, 33109954, 87169474, 265017, 265018, 265019, 264762, 18108351, 264763, 264683, 264388, 264288, 264685, 264768, 264687, 264768, 21906765, 21908788, 21908789, 55811857, 265020, 60431528, 263974, 18108379, 35695855, 264555, 264557, 264639, 83373044, 18108384, 87168578, 60432113, 22278000, 22279002, 264564, 264486
1892	95317232 (3983, 3984)	Novel Protein sim. GBank gij2246532 (U93872) - ORF 73, contains large complex repeat CR 73 (Kaposi's sarcoma-associated herpesvirus)	Contains protein domain (PF00096) - Zinc finger, C2H2 type	struct	29331827, 264906, 264807, 264809, 265007, 264603, 264766, 264688, 264768, 21906768, 264628, 264635, 264636, 18108385, 56528486, 264566, 264567
1893	80054783 (3985, 3986)	Novel Protein sim. GBank gij2565091 (U80761) - CTG26 alternate open reading frame [Homo sapiens]		UNCLASSIFIED	264592, 35698423

1884	84329114 (3987, 3988)	Novel Protein sim. GBank gi15630077 gb AA045822.1 AC006017 - (AC006017) similar to ALR; similar to AAC51735 (PID:g2358287) [Homo sapiens]	Contains protein domain (PF00856) - SET domain	mapolymerase	264488, 22278997, 22278998, 22278999, 29331822, 29331824, 29331825, 29331828, 29331827, 29331828, 264905, 264907, 264908, 265006, 60432229, 33657402, 60433356, 264757, 60434398, 264758, 33109954, 265011, 265017, 265018, 265019, 264684, 264368, 264685, 264686, 264768, 21906765, 21906767, 21906768, 21906769, 265020, 264690, 18108362, 264693, 65274620, 18108370, 264635, 264555, 264556, 264557, 58182323, 83373044, 56526486, 22279000, 22279002, 264564
1895	85414353 (3989, 3990)	Novel Protein sim. GBank gi14827040 ref NP_005110.1 pTRAP - thyroid hormone receptor-associated protein, 150 kDa subunit		UNCLASSIFIED	264488, 18108395, 22278994, 56984075, 22278998, 35696286, 22278997, 22278999, 264259, 29147620, 56182181, 29331824, 60432289, 29331826, 29331827, 35696052, 29146499, 264905, 264907, 66712502, 56182435, 265008, 265007, 265008, 265009, 60431735, 60433356, 33657402, 264595, 55812038, 33657084, 5581386, 85658542, 265010, 265011, 265017, 265018, 265019, 264761, 264762, 264448, 264683, 264764, 264288, 264766, 264686, 264768, 264769, 56181562, 264689, 21906765, 21906768, 21906767, 29148627, 21906768, 21906769, 29148629, 29148784, 265020, 265021, 264690, 18108361, 264693, 27486262, 27486264, 27486265, 18108370, 60431528, 18108374, 18108377, 35698423, 55811576, 65274791, 35695855, 264631, 264634, 264635, 264555, 264636, 60431850, 264557, 264558, 264559, 83373044, 20768451, 87188519, 264404, 80432113, 264587, 264564
1886	80254186 (3991, 3992)	Novel Protein sim. GBank gi1791146 emb CAA60020  - (X86028) extensin-like protein [Vigna unguiculata]		UNCLASSIFIED	
1897	87028423 (3993, 3994)	Novel Protein sim. GBank gi12642034 (AF034547) - protein phosphatase M130 myosin binding subunit [Ovis aries]	Contains protein domain (PF00023) - Ank repeat	phosphatase	264908, 264809, 284592, 284593
1898	85262704 (3995, 3996)	Novel Protein sim. GBank gi14589634 dbj BAAT6638.1  - (AB023212) KIAA0995 protein [Homo sapiens]		UNCLASSIFIED	264113, 264685, 264555, 264567



1899	94324903 (3997, 3998)	Novel Protein sim. GBank gil52253121gbJAA0848.1JAF07244 - (AF072441) calicheurin binding protein cabin 1 [Homo sapiens]	Contains protein domain (PF00515) - TPR Domain	UNCLASSIFIED	18108394, 18108397, 35696286, 60424269, 29331827, 29331828, 35696052, 265006, 264512, 55811386, 265010, 265018, 265019, 55811150, 18108351, 264763, 264682, 264389, 264685, 264688, 56181582, 265020, 264691, 33857023, 264693, 33857109, 27486284, 18108370, 18108378, 35695855, 264634, 264635, 264638, 264655, 264557, 56182323, 18108382, 264559, 83373044, 60432113, 22279000, 264563, 264564, 264566
2000	95413705 (3999, 4000)	Novel Protein sim. GBank gil1723232j9p1Q10151YATA_SCHPO - HYPOTHETICAL 90.6 KD PROTEIN C104.10 IN CHROMOSOME 1		UNCLASSIFIED	52646365, 52646842, 22278994, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 52845080, 29331822, 29331826, 29331827, 29331828, 35696052, 264106, 29331830, 52844045, 265007, 265008, 60170831, 264592, 264593, 33657402, 60433438, 21906754, 52644286, 265017, 265018, 265019, 264761, 264389, 264288, 52644228, 21908765, 21908768, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 60170815, 52644150, 33657023, 65274620, 52645129, 27486261, 27486262, 27486264, 35695763, 35696423, 35695855, 264631, 52844332, 56182323, 60170394, 83373044, 56526486, 22279002, 264566, 264567
2001	95072534 (4001, 4002)	Novel Protein sim. GBank gil107560ip1rj1B38637 - Ras inhibitor (clone JC265) - human (fragment)		UNCLASSIFIED	264769, 52644229, 65274572, 21906768, 22278996, 35696286, 35695917, 265020, 22278999, 264534, 264490, 264259, 264692, 60432289, 33657109, 35696052, 264508, 264509, 18108370, 60431528, 18108374, 35696423, 65274791, 35695855, 264510, 264511, 264512, 265009, 264634, 264636, 264555, 264556, 264638, 264557, 264558, 264559, 60433438, 83373044, 264759, 18108385, 265011, 264600, 264601, 60432113, 264603, 264604, 264605, 264448, 264288, 264765
2002	80236368 (4003, 4004)	Novel Protein sim. GBank gil729433p1P38657ER60_BOVIN - PROBABLE PROTEIN DISULFIDE ISOMERASE ER-60 PRECURSOR (ERP60) (58 KD MICROSOMAL PROTEIN) (P58)	Contains protein domain (PF00085) - Thioredoxin	isomerase	264907, 265006, 264910, 264603, 264692, 264829, 18108374, 264559, 264557
2003	80074449 (4005, 4006)	Novel Protein sim. GBank gil86388ip1rj1A27040 - neurofilament triplet M protein - chicken (fragment)		UNCLASSIFIED	264805, 264906, 264908, 264910, 264596, 265017, 18108351, 264692, 264629, 264634, 264555

2004	95317318 (4007, 4008)	Novel Protein sim. GBank gi 4884249 emb CAA3230.1  - (AL049996) hypothetical protein [Homo sapiens]	Contains protein domain (PF00076) - one ma_bind RNA recognition motif, (a.k.a. RRM, RBD, or RNP domain)	52645156, 52646842, 52646365, 56182575, 22278994, 22278995, 56994075, 22278999, 35696286, 22278997, 22278998, 22278999, 60432049, 264259, 52645080, 29331824, 29331826, 29331827, 35696052, 29331828, 33656870, 29331830, 264908, 264592, 60433356, 33657402, 52646317, 21906754, 33657084, 52644296, 87169474, 87168559, 265017, 265018, 265019, 264763, 264683, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 265022, 52644150, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 33857349, 27486265, 35695763, 18108374, 18108376, 18108377, 35696423, 35695855, 264631, 52644332, 264558, 18108385, 56526485, 87168518, 60432113, 264483, 264488, 264905
2005	87400864 (4009, 4010)	Novel Protein sim. GBank gi 3879501 emb CAA87795  - (Z47812) similar to ubiquitin carboxyl-terminal hydrolase; cDNA EST EMBL:D33368 comes from this gene; cDNA EST EMBL:D33965 comes from this gene; cDNA EST EMBL:D33822 comes from this gene; cDNA EST EMBL:D34547 comes from this gene...	ubiquitin	
2006	95351177 (4011, 4012)	Novel Protein sim. GBank gi 4106873 emb CAA22613  - (AL035064) queuine trna-ribosyltransferase [Schizosaccharomyces pombe]	Contains protein domain (PF01702) - Queuine trRNA-ribosyltransferase	56182575, 56181686, 22278995, 22278996, 22278997, 22278998, 22278999, 264259, 56182181, 60424288, 88714117, 35696052, 264606, 68712502, 264909, 264510, 60433356, 85658542, 265010, 265018, 265019, 264682, 264448, 264288, 264768, 29148627, 21906788, 29148784, 35695917, 60170615, 264691, 33657023, 65274620, 33657109, 55810764, 55811576, 35695855, 87168518, 60432113, 264583, 264482
2007	94325558 (4013, 4014)	Novel Protein sim. GBank gi 2662161 emb BAA23712  - (AB007900) HH0452 cDNA clone for KIAA0440 has a 438-bp insertion at position 1711 of the sequence of KIAA0440. [Homo sapiens]	UNCLASSIFIED	264488, 263894, 35696032, 264508, 264005, 264906, 264907, 264908, 264909, 265007, 264910, 264592, 264595, 264758, 265011, 264760, 264762, 264764, 264768, 264685, 264767, 264768, 264769, 55811957, 35695917, 265020, 264691, 264893, 264628, 264629, 65274791, 35695855, 264631, 264632, 264634, 264635, 264637, 264638, 264839, 264968
2008	85084428 (4015, 4016)	Novel Protein sim. GBank gi 1550783 emb CAA69257  - (Y07960) homeodomain protein [Mus musculus]	Contains protein domain (PF00046) - Homeobox domain	264909, 264768, 35695855







2035	83553451 (4089, 4070)				264369, 264686, 265022, 26526486, 264567
2036	87115833 (4071, 4072)				29331827, 29331828, 264682, 264369, 29148627, 60432113
2037	94324833 (4073, 4074)	Novel Protein sim. GBank gij2734081 (AF000195) - similar to oxysterol-binding proteins [Caenorhabditis elegans]		UNCLASSIFIED	65274572, 22278995, 22278996, 56994075, 35698286, 22278997, 22278998, 22278999, 264259, 29331824, 60432289, 29331828, 29331828, 35698052, 264807, 29331830, 66712502, 56182435, 265008, 265009, 60170831, 284594, 55812038, 33109954, 21906754, 87168559, 265017, 265018, 265019, 264762, 264369, 264288, 21906765, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 52644140, 33657023, 33657109, 33657182, 35695783, 35695855, 264632, 264634, 264636, 56182323, 83373044, 60432113, 22279000, 22279002, 264563
2038	95422384 (4075, 4076)	Novel Protein sim. GBank gij3880625[emb CAB07858] - (Z93785) predicted using GeneIndex; similar to RNA recognition motif (aka RRM, RBD, or RNP domain); cDNA EST EMBL: T01682 comes from this gene; cDNA EST EMBL: M75823 comes from this gene; cDNA EST EMBL: D27559 comes from this ge...	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	UNCLASSIFIED	22278995, 22278998, 56994075, 264259, 29331824, 35698052, 284905, 264906, 52644045, 265007, 265009, 87168559, 265017, 18108351, 284448, 264388, 264786, 264767, 264686, 18108358, 21906765, 21906768, 52644150, 33657023, 264692, 18108362, 33657109, 27485262, 18108370, 18108374, 18108378, 35698423, 65274791, 264632, 264636, 18108383, 83373044, 18108385, 87168518, 22279000, 22279002, 264563, 264584, 264586
2039	95514626 (4077, 4078)	Novel Protein sim. GBank gij224653[jdb BAA20813] - (AB00234) KIAA0356 [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	22278997, 264259, 29331822, 264905, 264906, 264907, 264908, 264909, 264510, 265009, 264910, 264593, 264758, 265011, 265018, 264762, 264288, 264768, 264768, 264769, 21906766, 33657023, 264692, 264693, 33657109, 35698423, 264631, 264632, 264634, 264635, 264636, 264637, 264639, 87168518, 264486
2040	95308417 (4079, 4080)			UNCLASSIFIED	264592
2041	95071736 (4081, 4082)	Novel Protein sim. GBank gij2500625[sp P70700 P2. MOUSE - DNA-DIRECTED RNA POLYMERASE I 135 KD POLYPEPTIDE (RNA POLYMERASE I SUBUNIT 2) (RPA135)	mapolymerase	UNCLASSIFIED	264488, 22278998, 35698052, 264905, 264907, 264908, 264910, 265018, 264605, 265019, 18108351, 264766, 264769, 21906766, 265021, 265022, 264692, 33657109, 264628, 264629, 35698423, 35695855, 264637, 264638, 264583, 264584, 264585, 264567

2042	95307447 (4083, 4084)	Novel Protein sim. GBank gij406590[gbjAAD20040] - (AF131766) Similar to Ena-VASP like protein [Homo sapiens]	Contains protein domain (PF00568) - WH1 domain	UNCLASSIFIED	60424178, 35696288, 284259, 29331826, 35696032, 29331828, 264508, 264509, 264907, 264909, 264510, 284511, 285009, 284910, 284591, 60433356, 284595, 285017, 265019, 264681, 264764, 264368, 264765, 264684, 264288, 264766, 264688, 52644229, 264769, 21908765, 35695817, 264535, 52644150, 264691, 264692, 18108365, 27486261, 27486262, 27486265, 18108374, 35698423, 65274781, 35695855, 284555, 264558, 60170394, 18108385, 264404, 22279000, 22279002, 264482, 264563, 264564, 264566
2043	94328076 (4085, 4086)	Novel Protein sim. GBank gij5042554[gbjAAD38607.1]AF145632) ScDNA.GH06032 (Drosophila melanogaster)	Contains protein domain (PF00122) - transport E1-E2 ATPase		264488, 52644507, 52646365, 56994075, 22278897, 22278898, 20281171, 264259, 29331822, 29331824, 86714117, 29331826, 29331828, 33656970, 29146498, 284509, 264908, 52644045, 56182435, 265006, 33657402, 21908754, 52644296, 87169559, 265017, 265018, 265019, 264681, 264288, 264766, 264685, 264686, 21908766, 21908767, 21908768, 21908769, 265020, 265021, 60170615, 264691, 33657023, 264693, 65274620, 33657109, 33657182, 27486281, 27486282, 33657349, 35695783, 18108374, 55811576, 35695855, 18108380, 18108381, 60170394, 56182323, 264558, 83373044, 18108385, 56526486, 87169518, 60432113, 22279000, 264567
2044	87106927 (4087, 4088)	Novel Protein sim. GBank gij2246532 (J93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]	Struct		264033, 29331827, 264905, 66712502, 264582, 264689, 21908765, 21908769, 265020, 264692, 264692, 264482, 264588
2045	79635532 (4089, 4090)				264692
2046	87320849 (4091, 4092)	Novel Protein sim. GBank gij406598[gbjAAD20062] - (AF131852) Unknown [Homo sapiens]			264259, 264906, 264683, 22279002
2047	84578801 (4093, 4094)	Novel Protein sim. GBank gij4101720 (AF006466) - lymphocyte specific formin related protein [Mus musculus]			22278999, 29147620, 29331824, 29146498, 264508, 265007, 265008, 265019, 264603, 264681, 29148827, 29148629, 265021, 33657023, 18108365, 33657109, 33657182, 18108377, 264558, 264638, 284559, 18108388
2048	84606378 (4095, 4096)			UNCLASSIFIED	264909
2049	88094690 (4097, 4098)	Novel Protein sim. GBank gij4589656[gbjBAA76850.1] - (AB023223) KIAA 1006 protein [Homo sapiens]		UNCLASSIFIED	264488, 264259, 29331824, 29331828, 35696032, 264906, 264907, 264908, 264909, 284910, 264603, 264763, 21806767, 21908768, 264629, 264634, 264637, 22279002, 264584, 264585, 284588, 284587

2050	79633835 (4089, 4100)			UNCLASSIFIED	264693	264488, 264259, 264509, 264906, 264907, 264769, 18108374, 35696423, 264563, 264566, 264488
2051	87780168 (4101, 4102)			UNCLASSIFIED		
2052	88096393 (4103, 4104)	Novel Protein sim. GBank gij4529889[gbjAAD21812.1] - (AF134726) G9A (Homo sapiens)	Contains protein domain (PF00856) - SET domain	kinase		264488, 263994, 35696032, 264508, 264905, 264508, 264906, 264907, 264908, 264909, 264113, 264511, 265009, 264910, 60170831, 264592, 264756, 265010, 265011, 264605, 264760, 264682, 264764, 264369, 264766, 264686, 264768, 264769, 52644229, 264689, 35695917, 33857023, 33857109, 264628, 18108374, 35696423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264556, 264638, 264639, 18108385, 56526486, 60432113, 264563, 264584, 264566, 264488, 264567, 264488, 263994, 35696032, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264113, 264511, 265009, 264910, 60170831, 264592, 264756, 265010, 265011, 264605, 264760, 264682, 264764, 264369, 264766, 264686, 264768, 264769, 52644229, 264689, 35695917, 33857023, 33857109, 264628, 18108374, 35696423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264556, 264638, 264639, 18108385, 56526486, 60432113, 264563, 264584, 264566, 264488, 264567
2053	87763078 (4105, 4106)	Novel Protein sim. GBank gij2995449[embjCAA75113] - (Y14848) midline 1 protein [Mus musculus]		UNCLASSIFIED		264488, 263994, 35696032, 264508, 264905, 264508, 264906, 264907, 264908, 264909, 264113, 264511, 265009, 264910, 60170831, 264592, 264756, 265010, 265011, 264605, 264760, 264682, 264764, 264369, 264766, 264686, 264768, 264769, 52644229, 264689, 35695917, 33857023, 33857109, 264628, 18108374, 35696423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264556, 264638, 264639, 18108385, 56526486, 60432113, 264563, 264584, 264566, 264488, 264567
2054	95358937 (4107, 4108)	Novel Protein sim. GBank gij3876326[embjCAB02090] - (Z79754) similar to C2 domain [Caenorhabditis elegans]	Contains protein domain (PF00168) - C2 domain			264488, 263994, 35696032, 264508, 264905, 264508, 264906, 264907, 264908, 264909, 264113, 264511, 265009, 264910, 60170831, 264592, 264756, 265010, 265011, 264605, 264760, 264682, 264764, 264369, 264766, 264686, 264768, 264769, 52644229, 264689, 35695917, 33857023, 33857109, 264628, 18108374, 35696423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264556, 264638, 264639, 18108385, 56526486, 60432113, 264563, 264584, 264566, 264488, 264567
2055	88259448 (4109, 4110)	Novel Protein sim. GBank gij5353746[gbjAAD42228.1]AF15913 - (AF159133) SIR2-like protein [Oryza sativa subsp. indica]		UNCLASSIFIED		264488, 263994, 35696032, 264508, 264905, 264508, 264906, 264907, 264908, 264909, 264113, 264511, 265009, 264910, 60170831, 264592, 264756, 265010, 265011, 264605, 264760, 264682, 264764, 264369, 264766, 264686, 264768, 264769, 52644229, 264689, 35695917, 33857023, 33857109, 264628, 18108374, 35696423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264556, 264638, 264639, 18108385, 56526486, 60432113, 264563, 264584, 264566, 264488, 264567



2056	88177396 (4111, 4112)	Novel Protein sim. GBank gi 4828860 ref NP_005042.1 pQARS - glutamine-IRNA synthetase	Contains protein domain (PF00749) - IRNA synthetases class I (E and Q)	synthase	284488, 52845156, 56182575, 22278994, 35995285, 56994075, 22278996, 22278998, 22278999, 60432049, 264259, 29331824, 60432289, 29331827, 29331828, 33656970, 264104, 264906, 284808, 265006, 265008, 60170831, 264591, 60432229, 6043438, 18108348, 21908754, 33657084, 52844296, 87168474, 265010, 87168559, 265017, 265018, 264760, 18108351, 264681, 264682, 264448, 264683, 284398, 264298, 264685, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906769, 55811957, 35995917, 265022, 33657023, 18108362, 33657109, 18108368, 33657182, 27486261, 27486264, 27486265, 33657349, 264628, 18108370, 264628, 18108374, 18108377, 18108379, 35898423, 55811578, 20281152, 264636, 284952, 18108385, 18108388, 87168518, 264482, 264565, 264566, 264567, 52646842, 52648365, 56182575, 35698286, 22278996, 22278997, 22278998, 264093, 52645080, 35698052, 29331828, 33656970, 265009, 52646317, 55811386, 52644296, 52644229, 21906769, 35695917, 265021, 60170615, 52644150, 33657109, 33657182, 27486261, 27486262, 35695763, 35696423, 35695855, 52644332
2057	87877905 (4113, 4114)	Novel Protein sim. GBank gi 728650 sp P08640 AMYH_YEAST - GLUCOAMYLASE S1/S2 PRECURSOR (GLUCAN 1,4-ALPHA- GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE)		UNCLASSIFIED	265007, 265008, 284591
2058	86276896 (4115, 4116)	Novel Protein sim. GBank		UNCLASSIFIED	29331825, 264682, 264686, 264691, 264693, 22279002
2059	79866684 (4117, 4118)	gi 119714 sp P13983 EXTN_TOBAC - EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)		UNCLASSIFIED	
2060	83050800 (4119, 4120)	Novel Protein sim. GBank gi 281122 (U87318) - NaDC-2 (Xenopus laevis)		UNCLASSIFIED	56182575, 29331824, 29331826, 2646910, 55811957, 18108370, 55811576

2061	85362204 (4121, 4122)	Novel Protein sim. GBank gi 2496947 sp Q09298 YQ09_CAEEL - HYPOTHETICAL 141.2 KD PROTEIN EEED8.9 IN CHROMOSOME II	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain	22278997, 22278999, 264259, 29331822, 29331824, 29331826, 29331827, 3569052, 29331828, 264908, 66712502, 29331830, 264908, 264909, 284112, 284511, 265007, 265008, 264910, 264591, 33657402, 21906754, 85658542, 265017, 265019, 264448, 264683, 264288, 264684, 264369, 264686, 264687, 18108358, 264689, 21906765, 21906767, 21906768, 21906769, 265020, 265022, 264691, 33657023, 33657109, 20281149, 18108379, 35695855, 264634, 264556, 264557, 284558, 18108382, 284559, 83373044, 18108384, 56528488, 60432113
2062	87028440 (4123, 4124)	Novel Protein sim. GBank gi 4502091 ref NP_001139.1 pANK2 - ankyrin 2, neuronal	Contains protein domain (PF00023) - struct Ank repeat	264905, 264628, 264907, 264629, 264908, 264909, 18108374, 263978, 35695855, 264512, 264635, 60431850, 264636, 264760, 264563, 18108351, 264782, 264565, 264784, 264487, 264766
2063	87601272 (4125, 4126)	Novel Protein sim. GBank gi 4589562 dbj BAA76803.1  - (AB023176) KIAA0959 protein [Homo sapiens]	Contains protein domain (PF00617) - oncogene RasGEF domain	22278994, 22278999, 264259, 29331827, 264906, 264909, 52644045, 264686, 21906767, 55811957, 264692, 18108365, 263972, 55811576, 18108384, 22279002, 264482, 264563, 264564, 264484
2064	95317253 (4127, 4128)	Novel Protein sim. GBank gi 1754515 dbj BAA13413.1  - (D87515) aminopeptidase-B [Rattus norvegicus]	hydrolase	264488, 52646365, 56994075, 35696286, 22278997, 22278998, 264259, 29331828, 60432289, 29331827, 29331828, 35696052, 264509, 265007, 265008, 60432229, 60433438, 21906754, 265010, 265011, 87168559, 265017, 265018, 264761, 18108351, 264682, 264369, 264288, 52644229, 21906765, 21906767, 21906768, 35695917, 33657109, 18108368, 18108374, 35696423, 35695855, 52644332, 264559, 60432113, 22279000, 22279002, 264566, 264486
2065	95092238 (4129, 4130)	Novel Protein sim. GBank gi 2507144 sp Q04203 TENS_CHICK - TENSIN	kinase	264569, 18108394, 56182181, 60432289, 29331826, 264905, 264906, 264908, 60431735, 60433356, 55811386, 85658542, 265018, 55811150, 264681, 264766, 264692, 60431528, 263974, 55810764, 35695855, 264631, 264634, 264635, 60431850, 264557, 83373044, 18108388, 22279000, 22279002, 56182575, 264259, 264906, 264764, 264288, 56182323, 264567
2066	85783402 (4131, 4132)	Novel Protein sim. GBank gi 160171 (M58295) - circumsporozoite protein [Plasmodium yoelii]	Contains protein domain (PF00096) - UNCLASSIFIED Zinc finger, C2H2 type	

2067	85303892 (4133, 4134)				35698286, 22278997, 22278998, 60432049, 264259, 60432289, 60433438, 264682, 264448, 264369, 264288, 18108355, 21906765, 21906768, 265022, 33657109, 35698423, 35698555, 264558, 264404, 264583, 264486
2068	84344754 (4135, 4136)			UNCLASSIFIED	264687
2069	84318177 (4137, 4138)	Novel Protein sim. GBank gij3152682 (AF084604) - KE03 protein [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	transcript factor	60424179, 56182575, 22278995, 22278996, 56994075, 264259, 29331822, 29331824, 29331825, 35698052, 29331826, 33656970, 264508, 264805, 56182435, 265008, 60433356, 87168559, 285017, 265018, 264604, 265019, 264448, 264764, 264766, 21906765, 21906767, 21906768, 21906769, 265020, 265021, 33657023, 33657109, 263976, 264555, 264557, 56182323, 83373044, 87168518, 60432113, 22278900, 22278902
2070	85791380 (4139, 4140)	Novel Protein sim. GBank gij5712131gbjAAD47379.1 AF12049 - (AF120499) DEM1 protein [Homo sapiens]		UNCLASSIFIED	35595917, 264905, 264628, 264908, 264638
2071	86946116 (4141, 4142)	Novel Protein sim. GBank gij3551531 dbj BAA33016  - (AB017437) avena [Gallus gallus]	Contains protein domain (PF00568) - WH1 domain		18108398, 265008, 265007, 265008, 265009, 264584, 265010, 265011, 18108351, 18108354, 18108364, 18108365, 18108368, 264634, 18108381, 18108385, 18108386, 18108391
2072	91718429 (4143, 4144)		Contains protein domain (PF00184) - Neurohypophyseal hormones, C-terminal Domain		22278995, 35698286, 22278997, 22278998, 22278999, 264490, 60432049, 264259, 29331822, 29331824, 29331826, 35698052, 265008, 33657402, 21906754, 265011, 265018, 18108351, 264682, 264369, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 264690, 264692, 35698423, 264555, 264556, 264558, 22279000
2073	27825664 (4145, 4146)	Novel Protein sim. GBank gij1504026 dbj BAA13212  - (D86976) similar to C.elegans protein (Z37093) [Homo sapiens]		UNCLASSIFIED	264556
2074	94324767 (4147, 4148)	Novel Protein sim. GBank gij4240317 dbj BAA74937.1  - (AB020721) KIAA0914 protein [Homo sapiens]			29331822, 264909, 264511, 265008, 264594, 264595, 265010, 265011, 265017, 265018, 265019, 264448, 264683, 265020, 265021, 18108370, 264632, 83373044, 264567

2075	94314886 (4149, 4150)	Novel Protein sim. GBank gi 5138930 gb AA040382.1  - (AF093680) transcription factor IIB [Homo sapiens]			18108384, 22278994, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331823, 29331827, 35898052, 29331828, 264905, 264907, 264908, 264510, 265007, 264910, 265009, 33657084, 264760, 264448, 264288, 264766, 264767, 264689, 21906786, 21906787, 21906789, 265021, 285022, 60170615, 33657023, 27486282, 27486265, 35896423, 35895855, 264631, 264634, 264639, 87168518, 22279002, 264563, 264486, 18108391
2076	87594118 (4151, 4152)				264259, 29331828, 264508, 264908, 264510, 265007, 285011, 284288, 284637, 18108385, 284592
2077	11389877 (4153, 4154)			UNCLASSIFIED	
2078	87539384 (4155, 4156)	Novel Protein sim. GBank gi 420590 db BAA74578  - (D87908) nuclear protein np85 [Mus musculus]	Contains protein domain (PF00628) - PHD-finger	ubiquitin	29331825, 265017, 265018, 264288, 265020, 265021, 264634, 56528486
2079	88095916 (4157, 4158)	Novel Protein sim. GBank gi 424025 db BAA74906.1  - (AB020690) KIAA0883 protein [Homo sapiens]	Contains protein domain (PF000098) - Zinc finger, CCHC class	UNCLASSIFIED	264768, 22278997, 265021, 264690, 264259, 264692, 29331822, 264693, 29331824, 29331828, 264508, 264509, 264908, 264907, 264628, 20281069, 264909, 265007, 265009, 264632, 264636, 264591, 264592, 264639, 264758, 264759, 33108954, 264604, 265018, 285019, 22278002, 284563, 284564, 264448, 264684, 264567, 264685
2080	94136689 (4159, 4160)	Novel Protein sim. GBank gi 240802 emb CA816219.1  - (Z98162) putative vacuolar protein [Schizosaccharomyces pombe]	Contains protein domain (PF01363) - FYVE zinc finger	UNCLASSIFIED	56182575, 264092, 29331824, 29331826, 29331830, 265017, 265018, 265020, 83373044
2081	94847186 (4161, 4162)	Novel Protein sim. GBank gi 5524734 gb AA044360.1 AF16835 - (AF168350) ST7 protein [Homo sapiens]	Contains protein domain (PF00431) - CUB domain	eph	22278998, 22278997, 22278999, 60432049, 29331822, 29331824, 33857402, 85658542, 265011, 265018, 265019, 21908787, 21908788, 264683, 18108385, 22278000, 22279002
2082	87628629 (4163, 4164)	Novel Protein sim. GBank gi 3880558 emb CAA94234  - (Z70271) predicted using GeneFinder; similar to collagen; cDNA EST yk308e7.3 comes from this gene; cDNA EST yk308e7.5 comes from this gene; cDNA EST yk385a8.3 comes from this gene; cDNA EST yk385a8.5 comes from this gene [Caeno...]		collagen	264907, 265019
2083	94141000 (4165, 4166)	Novel Protein sim. GBank gi 2352427 AF004161  - peroxisomal Ca-dependent solute carrier [Oryctolagus cuniculus]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	56182575, 22278996, 22278997, 264259, 29331822, 29331825, 264509, 264112, 265009, 264593, 21906754, 265018, 265019, 264448, 264288, 264685, 21906769, 265020, 265022, 284691, 18108370, 85274791, 264631, 284555, 284583
2084	95199298 (4167, 4168)	Novel Protein sim. GBank gi 728836 sp P39193 ALU6_HUMAN . III  ALU SUBFAMILY SP WARNING ENTRY III		UNCLASSIFIED	265018, 264763, 264683, 284691

2085	94988478 (4169, 4170)	Novel Protein sim. GBank gij1655609[emb CAA6032]- (Y07752) perophorin-S [voiox carten]	UNCLASSIFIED	56182575, 60432289, 284908, 56182435, 87168474, 264763, 264389, 264688, 264693, 18108370, 56182323
2086	91234404 (4171, 4172)	Novel Protein sim. GBank gij3875032[emb CAA88936] - [Z49125] similarity to Trichostrongylus colubriformis 11 kd secretory protein (Swiss Prot accession number P21937); cDNA EST EMBL:D33349 comes from this gene; cDNA EST EMBL:D37644 comes from this gene; cDNA EST EMBL:D36149 come...	UNCLASSIFIED	35696286, 264259, 35986052, 264906, 264807, 264908, 264909, 264910, 264759, 264604, 264762, 264768, 264769, 35695917, 263978, 35698423, 35695855, 264632, 264634, 264637, 264638, 264639, 56182323, 18108385, 264482, 264486
2087	21436337 (4173, 4174)		UNCLASSIFIED	264489
2088	94111527 (4175, 4176)	Novel Protein sim. GBank gij3880930[emb CAA16334.1] - (AL021481) similar to Phosphoglucosylase and phosphomannomylase phosphoserine; cDNA EST EMBL:D36168 comes from this gene; cDNA EST EMBL:D70697 comes from this gene; cDNA EST yk373h9.5 comes from this gene; cDNA EST EMBL:T0080...		264488, 22278994, 35696286, 22278996, 29331827, 35696052, 33657402, 21906754, 33109954, 87168474, 265017, 265018, 265019, 264448, 264683, 264389, 264685, 264687, 264689, 21908765, 21906766, 21906767, 21906768, 21908769, 265020, 265021, 265022, 264682, 33657023, 33657109, 33657182, 27486261, 27488262, 33657349, 27486265, 35698423, 35695855, 83373044, 87168518, 22279000, 264587
2089	85422801 (4177, 4178)	Novel Protein sim. GBank gij4758118[ref NP_004623.1]pDAP3 - Death associated protein 3	cadherin	18108392, 264488, 52644507, 18108394, 18108397, 52646842, 18108398, 56182575, 22278994, 22278995, 35698286, 22278996, 56994075, 22278997, 22278998, 22278999, 264091, 264092, 264093, 264094, 60432049, 264259, 29331822, 20281099, 29331824, 29331825, 29331826, 29331827, 29331828, 35698052, 33656970, 29146498, 29146499, 264102, 264106, 264107, 264109, 264508, 264905, 264509, 264906, 264907, 264908, 66712502, 264828, 52644045, 264809, 56182435, 264110, 264112, 264510, 264511, 265006, 264512, 265007, 265008, 264910, 265009, 60170831, 264592, 264593, 80433356, 33657402, 60433438, 264595, 55812038, 264758, 21906754, 33657084, 55811388, 52644298, 265010, 265011, 87168559, 265017, 265018, 265019, 264760, 264761, 55811150, 264762, 18108351, 264682, 264448, 264763, 264764, 264683, 264389, 18108354, 264288, 264685, 264768, 264686, 264687, 264768, 52644228, 264688, 18108358, 56181562, 264769, 18108359, 264688, 21906765, 21906766, 21906767, 21906768, 29148627, 21906769, 55811857, 29148629, 29148784, 35695917, 265020, 265021, 265022, 60170815, 264680,

2090	88222470 (4179, 4180)				22278995, 22278998, 22278999, 264259, 29331826, 35696052, 264910, 33657402, 60433438, 33109954, 87168474, 87168359, 265018, 265019, 264681, 264684, 264686, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906768, 35696423, 265022, 60170615, 33657023, 35696423, 35695855, 264952, 18108387, 22279000, 263984, 264905, 264908, 264511, 264512, 265008, 264910, 55811386, 264288, 264768, 56181562, 21906765, 21906768, 21906769, 265022, 264628, 264563, 264557, 22278997, 22278999, 66712502, 87168559, 264683, 265021, 264486
2091	95309181 (4181, 4182)	Novel Protein sim. GBank gij4560997(gbAAD24571.1)AF12108 - (AF121081) cAMP inducible 2 protein [Mus musculus]		UNCLASSIFIED	22278995, 22278998, 22278999, 264259, 29331826, 35696052, 264910, 33657402, 60433438, 33109954, 87168474, 87168359, 265018, 265019, 264681, 264684, 264686, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906768, 35696423, 265022, 60170615, 33657023, 35696423, 35695855, 264952, 18108387, 22279000, 263984, 264905, 264908, 264511, 264512, 265008, 264910, 55811386, 264288, 264768, 56181562, 21906765, 21906768, 21906769, 265022, 264628, 264563, 264557, 22278997, 22278999, 66712502, 87168559, 264683, 265021, 264486
2092	88223605 (4183, 4184)			homeobox	22278995, 22278998, 22278999, 264259, 29331826, 35696052, 264910, 33657402, 60433438, 33109954, 87168474, 87168359, 265018, 265019, 264681, 264684, 264686, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906768, 35696423, 265022, 60170615, 33657023, 35696423, 35695855, 264952, 18108387, 22279000, 263984, 264905, 264908, 264511, 264512, 265008, 264910, 55811386, 264288, 264768, 56181562, 21906765, 21906768, 21906769, 265022, 264628, 264563, 264557, 22278997, 22278999, 66712502, 87168559, 264683, 265021, 264486
2093	87406073 (4185, 4186)	Novel Protein sim. GBank gij2352427 (AF004161) - peroxisomal Ca-dependent solute carrier (Oryctolagus cuniculus)	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	22278995, 22278998, 22278999, 264259, 29331826, 35696052, 264910, 33657402, 60433438, 33109954, 87168474, 87168359, 265018, 265019, 264681, 264684, 264686, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906768, 35696423, 265022, 60170615, 33657023, 35696423, 35695855, 264952, 18108387, 22279000, 263984, 264905, 264908, 264511, 264512, 265008, 264910, 55811386, 264288, 264768, 56181562, 21906765, 21906768, 21906769, 265022, 264628, 264563, 264557, 22278997, 22278999, 66712502, 87168559, 264683, 265021, 264486
2094	91230929 (4187, 4188)	Novel Protein sim. GBank gij4929551(gbAAD34036.1)AF15179 - (AF151799) CGI-40 protein [Homo sapiens]		MHC	22278995, 22278998, 22278999, 264259, 29331826, 35696052, 264910, 33657402, 60433438, 33109954, 87168474, 87168359, 265018, 265019, 264681, 264684, 264686, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906768, 35696423, 265022, 60170615, 33657023, 35696423, 35695855, 264952, 18108387, 22279000, 263984, 264905, 264908, 264511, 264512, 265008, 264910, 55811386, 264288, 264768, 56181562, 21906765, 21906768, 21906769, 265022, 264628, 264563, 264557, 22278997, 22278999, 66712502, 87168559, 264683, 265021, 264486
2095	95351526 (4189, 4190)	Novel Protein sim. GBank gij1383238(pir)A57284 - spermatid perinuclear RNA-binding protein Spnr - mouse	Contains protein domain (PF000035) - Double-stranded RNA binding motif	dna_ma_bind	22278995, 22278998, 22278999, 264259, 29331826, 35696052, 264910, 33657402, 60433438, 33109954, 87168474, 87168359, 265018, 265019, 264681, 264684, 264686, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906768, 35696423, 265022, 60170615, 33657023, 35696423, 35695855, 264952, 18108387, 22279000, 263984, 264905, 264908, 264511, 264512, 265008, 264910, 55811386, 264288, 264768, 56181562, 21906765, 21906768, 21906769, 265022, 264628, 264563, 264557, 22278997, 22278999, 66712502, 87168559, 264683, 265021, 264486
2096	94119760 (4191, 4192)	Novel Protein sim. GBank gij3534423 (AF070689) - cytoplasmic dynein intermediate chain isoform DIC1a [Drosophila melanogaster]	Contains protein domain (PF00400) - WD domain, G-beta repeat	ATPase_associated	22278995, 22278998, 22278999, 264259, 29331826, 35696052, 264910, 33657402, 60433438, 33109954, 87168474, 87168359, 265018, 265019, 264681, 264684, 264686, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906768, 35696423, 265022, 60170615, 33657023, 35696423, 35695855, 264952, 18108387, 22279000, 263984, 264905, 264908, 264511, 264512, 265008, 264910, 55811386, 264288, 264768, 56181562, 21906765, 21906768, 21906769, 265022, 264628, 264563, 264557, 22278997, 22278999, 66712502, 87168559, 264683, 265021, 264486

2097	95322772 (4193, 4194)	Novel Protein sim. GBank g 5174501 re NP_006051.1 pLYF1 - zinc finger protein, subfamily 1A, 1 (Ikaros)	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	65274572, 264511, 265010, 264600, 265017, 264448, 264288, 265021, 60170615, 264692, 33657109, 18108370, 264636, 264483, 56984075, 284259, 264288, 265020, 264563
2098	87780340 (4195, 4196)	Novel Protein sim. GBank g 4758208 re NP_004081.1 pDUSP - dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	Contains protein domain (PF00782) - Dual specificity phosphatase, catalytic domain	phosphatase	65274572, 264905, 65274444, 264691, 264636, 264555
2099	95412927 (4197, 4198)	Novel Protein sim. GBank g 2695659 (AF028954) - pyruvate dehydrogenase phosphatase regulatory subunit precursor, PDP1 (Bos taurus)	Contains protein domain (PF00025) - ADP-ribosylation factor family	phosphatase	56182575, 22278995, 22278996, 22278997, 22278998, 60432049, 264259, 29331822, 29331824, 29331825, 29331827, 29331828, 29146486, 264909, 265008, 265009, 264910, 284591, 60432229, 60433356, 33657402, 284758, 21906754, 85658542, 87168474, 285017, 285018, 265019, 264681, 18108351, 264762, 264448, 264369, 264288, 18108354, 264686, 21906755, 21906767, 21906788, 21906769, 265020, 285021, 33657023, 18108374, 35696423, 264558, 83373044, 87168518, 60432113, 22279000, 22279002, 284091, 29331824, 264105, 265007, 265010, 18108380
2101	87762604 (4201, 4202)	Novel Protein sim. GBank g 4589468 db BAA76761.1  - (AB012808) mBOCT [Mus musculus]	Contains protein domain (PF01530) - Zinc finger, C2HC type	UNCLASSIFIED	264488, 264489, 35696286, 264259, 35696052, 264508, 264905, 264907, 264908, 264909, 264511, 264512, 264591, 264593, 60433356, 264758, 264601, 264605, 264760, 18108351, 264448, 264764, 264288, 264767, 264768, 21906769, 35695917, 18108374, 264634, 264555, 264559, 264563, 264482, 264486
2102	87770461 (4203, 4204)	Novel Protein sim. GBank g 3874149 emb CAA97423.1  - (Z73103) predicted using GeneFinder (Caenorhabditis elegans)	Contains protein domain (PF01530) - Zinc finger, C2HC type	UNCLASSIFIED	65274572, 56984075, 22278999, 264259, 29331824, 29331825, 35696052, 29331828, 66712502, 285009, 60170831, 284595, 33109954, 85658542, 87168559, 265017, 265019, 264448, 21906765, 21906768, 285022, 33657023, 27486262, 33657349, 35695763, 60431528, 18108374, 55811576, 56182323, 18108387, 87168518, 60432113, 264592, 264604, 22279000
2103	95413576 (4205, 4206)	Novel Protein sim. GBank g 4240159 db BAA74858.1  - (AB020642) KIAA0835 protein (Homo sapiens)	Contains protein domain (PF01530) - Zinc finger, C2HC type	transcriptfactor	65274572, 56984075, 22278999, 264259, 29331824, 29331825, 35696052, 29331828, 66712502, 285009, 60170831, 284595, 33109954, 85658542, 87168559, 265017, 265019, 264448, 21906765, 21906768, 285022, 33657023, 27486262, 33657349, 35695763, 60431528, 18108374, 55811576, 56182323, 18108387, 87168518, 60432113, 264592, 264604, 22279000
2104	85776161 (4207, 4208)			UNCLASSIFIED	

2105	94848080 (4209, 4210)	Novel Protein sim. GBank gji1707032 (U80445) - coded for by C. elegans cDNA yk13g5.3; coded for by C. elegans cDNA yk21g6.3; coded for by C. elegans cDNA CEMSE18F; coded for by C. elegans cDNA yk12b1.3; coded for by C. elegans cDNA yk65h8.3; coded for by C. elegans cDNA yk65h8.....		UNCLASSIFIED	264488, 56182575, 22278994, 56994075, 22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331826, 60432289, 29331827, 29331828, 264508, 264905, 264509, 264907, 29331830, 52644045, 264510, 264511, 265007, 264512, 265009, 60170831, 60432229, 33857402, 60433356, 264595, 60433438, 264758, 33857084, 87168474, 265010, 87168559, 265017, 265018, 265019, 264782, 18108351, 264684, 18108354, 264288, 264688, 52644228, 18108359, 21806765, 21906768, 21906767, 21908768, 21908769, 35695917, 265020, 265022, 60170615, 52644150, 264692, 33857023, 264693, 33857109, 60431528, 18108374, 85274781, 35695855, 264635, 60170394, 264639, 264558, 18108385, 18108387, 56526486, 87168518, 60432113, 264584, 264586, 264587, 265006, 265019
2106	83365475 (4211, 4212)	Novel Protein sim. GBank gji3881524[emb CAA93883] - (Z70038) ZK1067.4 [Caenorhabditis elegans]			264908, 264639
2107	79822662 (4213, 4214)	Novel Protein sim. GBank gji3176889 (AC003671) - Contains similarity to ubiquitin carboxyl-terminal hydrolase 14 gp[Z35927 from S. cerevisiae, [Arabidopsis thaliana]		UNCLASSIFIED	264905, 264906, 264907, 264908, 264909, 264758, 265011, 264600, 264601, 264764, 264766, 264767, 264768, 264769, 264693, 264629, 35695855, 264632, 264634, 264635, 264638, 264639, 83373044, 264486
2108	84233376 (4215, 4216)			UNCLASSIFIED	18108348, 264769, 18108370, 18108374, 284555, 264556, 264557, 284558
2109	80478719 (4217, 4218)	Novel Protein sim. GBank gji481043[pir J537671] - bat2 protein - human		UNCLASSIFIED	284564
2110	87729075 (4219, 4220)	Novel Protein sim. GBank gji2143639[pir J56542] - calmodulin-binding protein - rat	Contains protein domain (PF00069) - struct	UNCLASSIFIED	264766, 35695917, 264630, 264587, 264486
2111	87818419 (4221, 4222)	Novel Protein sim. GBank gji4426629[gb AAD20459] - (AF100960) protocadherin [Rattus norvegicus]	Eukaryotic protein kinase domain		264508, 264906, 264591, 264682, 22279002
2112	87293783 (4223, 4224)	Novel Protein sim. GBank gji3327184[dbj BAA31660] - (AB014585) KIAA0685 protein [Homo sapiens]	Contains protein domain (PF00028) - cadherin		265006
2113	78841388 (4225, 4226)		Cadherin domain		56714117, 29331828, 29331827, 60433438, 55812038, 265017, 265019, 264889, 21906768, 55811957, 265020, 265021, 33557109, 60170394, 264558
2114	87889342 (4227, 4228)	Novel Protein sim. GBank gji4757890[re NP_004328.1]pC8OR - chromosome 8 open reading frame 1		UNCLASSIFIED	65274572, 264689, 264691, 264692, 60432113
2115	90993785 (4229, 4230)	Novel Protein sim. GBank gji4757890[re NP_004328.1]pC8OR - chromosome 8 open reading frame 1		UNCLASSIFIED	



2116	88259387 (4231, 4232)	Novel Protein sim. GBank gi 2245532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		struct	35696286, 22278999, 56182181, 29331824, 29331825, 29331827, 35696052, 284907, 56182435, 265008, 264591, 55812038, 55811386, 87188559, 264288, 264369, 21906758, 29148528, 33657023, 35695763, 55811576, 35696423, 18108385
2117	87788904 (4233, 4234)	Novel Protein sim. GBank gi 2330021 (AF019250) - Kinetin-related protein; KRP; Costal2 [Drosophila melanogaster]		struct	28331824, 264511, 265009, 33109954, 265017, 265018, 264288, 264689, 265020, 264692, 56526488, 264482
2118	87078694 (4235, 4236)	Novel Protein sim. GBank gi 1078307 pir B56573 - nuclear pore complex glycoprotein p62 - African clawed frog		glycoprotein	264259, 264905, 264907, 264908, 264510, 264511, 265009, 264910, 265010, 264602, 264288, 264768, 264693, 263967, 263972, 264638, 264559
2119	86988317 (4237, 4238)	Novel Protein sim. GBank gi 4321407 gb AAD15748 - (AF047690) ATP-binding cassette protein M-ABC1 [Homo sapiens]		UNCLASSIFIED	264693, 18108365
2120	87788395 (4239, 4240)	Novel Protein sim. GBank gi 4865527 ref NP_005480.1 pNSP3 - novel SH2-containing protein 3	Contains protein domain (PF00017) - eph Src homology domain 2		264091, 264259, 29331826, 29331828, 265017, 264604, 264288, 264685, 265020, 264691, 18108370, 55810764, 264555, 264636, 60432113
2121	80021375 (4241, 4242)	Novel Protein sim. GBank gi 4757728 ref NP_004886.1 pAGTA - angiotensin/Vasopressin receptor ALLAVP-like		UNCLASSIFIED	264601, 264766, 263978
2122	91230931 (4243, 4244)	Novel Protein sim. GBank gi 4929551 gb AAD34036.1 AF15179 - (AF151799) CGI-40 protein [Homo sapiens]			18108394, 56182575, 22278697, 29331822, 29331824, 29331825, 29331826, 29331828, 264907, 56182435, 265007, 264910, 265010, 265018, 264686, 265020, 55811576, 264555, 264637, 18108382, 83373044, 18108383, 18108384, 56526486, 264565, 264567
2123	86787998 (4245, 4246)	Novel Protein sim. GBank gi 2224551 dbj BAA20764 - (AB002303) KIAA0305 [Homo sapiens]	Contains protein domain (PF01363) - F-YVE zinc finger	struct	18108396, 264757, 265011, 18108351, 264691, 264634, 18108385
2124	83005951 (4247, 4248)	Novel Protein sim. GBank gi 5689455 dbj BAA83011.1 - (AB028982) KIAA1059 protein [Homo sapiens]	Contains protein domain (PF00801) - PKD domain	transport	29331822, 264906, 264907, 264591, 264639, 264593
2125	95354041 (4249, 4250)	Novel Protein sim. GBank gi 728831 sp P39186 ALU1_HUMAN - III ALU SUBFAMILY J WARNING ENTRY IIII		UNCLASSIFIED	284259, 284509, 284907, 284511, 85658542, 284763, 21906785, 35695917, 264638, 264488
2126	95084231 (4251, 4252)	Novel Protein sim. GBank gi 4539264 emb CAB39853.1 - (AL049495) conserved hypothetical protein [Schizosaccharomyces pombe]		UNCLASSIFIED	264488, 264489, 29331827, 35696052, 264905, 264509, 264908, 264909, 264510, 265009, 264591, 264592, 264593, 33657402, 264594, 264595, 264596, 264758, 264601, 264603, 285018, 264604, 264605, 264760, 264681, 264762, 264683, 264764, 264684, 264288, 264685, 264689, 80170615, 33657023, 33657109, 55810764, 264635, 264636, 264637, 264638, 264639, 83373044, 264554, 264568

2127	81118652 (4253, 4254)	Novel Protein sim. GBank gi 4868435 gb AAD31315.1 AF14323 - (AF143236) apoptosis related protein APR-2 (Homo sapiens)			35696286, 29331828, 35699052, 284508, 284509, 284905, 284906, 284907, 284908, 284909, 284510, 285006, 284511, 284512, 285007, 285009, 284910, 284758, 285011, 284600, 284601, 284604, 284762, 284763, 284766, 284687, 284768, 284769, 284689, 35695917, 284690, 284691, 284692, 284693, 284629, 18108374, 35695855, 284632, 284634, 284635, 284636, 284637, 284638, 284639, 18108385, 284583, 284584, 284587, 56182575, 35696286, 56182181, 29331824, 60432289, 35696052, 284905, 284907, 66712502, 284908, 284909, 284510, 284512, 285009, 284910, 284591, 55812038, 285018, 284764, 284288, 284368, 284687, 284768, 55811957, 284892, 18108386, 284628, 284632, 284634, 284635, 284637, 56182323, 284639, 18108384, 18108388, 284563, 284567
2128	87414282 (4255, 4256)				66714117, 284828, 284595, 55812038, 55811150, 55811957, 284683, 18108374, 283978, 65274791, 18108381, 83373044, 22279000
2129	95102089 (4257, 4258)				UNCLASSIFIED
2130	95417144 (4259, 4260)	Novel Protein sim. GBank gi 2649255 (AE001012) - conserved hypothetical protein [Archaeoglobus fulgidus]			UNCLASSIFIED
2131	85723065 (4261, 4262)	Novel Protein sim. GBank gi 1085986 (U41276) - Similar to potassium channel protein, [Caenorhabditis elegans]	Contains protein domain (PF00805) - Pentapeptide repeats (8 copies)		35696052, 284909, 284768, 35695917
2132	85361096 (4263, 4264)	Novel Protein sim. GBank gi 5689373 db BAA82973.1 - (AB028844) KIAA1021 protein [Homo sapiens]	Contains protein domain (PF00122) - E1-E2 ATPase		284488, 22278999, 284259, 29331827, 29331828, 35696052, 284509, 284905, 284906, 284907, 284908, 284909, 284510, 285006, 284511, 285007, 284910, 284591, 284592, 284595, 284758, 21808754, 33109954, 87168474, 265011, 284600, 284601, 284605, 265018, 284760, 18108351, 284681, 284782, 284784, 284288, 284684, 284766, 284686, 284687, 284768, 284769, 284688, 21808769, 284690, 52844150, 284691, 284693, 18108370, 284628, 284629, 18108372, 18108374, 35696423, 35695855, 284631, 284634, 284635, 284636, 284555, 284637, 18108380, 284639, 284556, 56182323, 56525486, 284564, 284565, 284566, 284567

2133	95351538 (4265, 4266)	Novel Protein sim. GBank gi 4220489 (AC006069) - hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	50424170, 52646355, 52646842, 56994075, 3598288, 22278897, 22278898, 60432049, 56182181, 66714117, 60424268, 29331826, 29331828, 35696032, 264905, 264906, 264907, 66712502, 29331830, 56182435, 265006, 264512, 265008, 60431735, 60433358, 33657402, 55812038, 33109854, 21908754, 55811385, 265010, 264603, 265017, 265018, 265019, 55811150, 18108351, 264682, 284368, 284288, 52644229, 56181582, 21908765, 21908766, 21908767, 21908768, 21908769, 55811957, 35695917, 265020, 265021, 60170815, 33657023, 33657109, 60431528, 18108374, 35696423, 65274791, 35695855, 264634, 60431850, 18108380, 56182323, 83373044, 18108385, 18108387, 60432113, 22279002, 264563, 264566
2134	95412697 (4267, 4268)	Novel Protein sim. GBank gi 3875351(emb CA08415) - (Z96047) DY3.6 [Caenorhabditis elegans]			56181686, 35696286, 21906754, 55811386, 265011, 265017, 18108351, 264765, 264766, 264888, 21908768, 35695917, 265020, 33657023, 264628, 35695855, 264632, 264555, 264556, 264557, 264558, 18108382, 22279002
2135	88079813 (4269, 4270)	Novel Protein sim. GBank gi 5689559(emb BAA83063.1) - (AB029034) KIAA1111 protein [Homo sapiens]	Contains protein domain (PF00628) - PHD-finger	UNCLASSIFIED	22278999, 29331828, 35696052, 264906, 264908, 264910, 265009, 264591, 264758, 52646317, 265011, 87168559, 264601, 18108351, 264448, 264683, 264684, 264689, 18108359, 264691, 33657023, 264692, 35695763, 264629, 35695855, 264631, 264635, 264636, 264637, 56182323, 264639, 22278002, 264564
2136	84346478 (4271, 4272)	Novel Protein sim. GBank gi 2662167(emb BAA23715) - (AB007903) KIAA0443 [Homo sapiens]		UNCLASSIFIED	264539
2137	87637716 (4273, 4274)	Novel Protein sim. GBank gi 4884110(emb CAB43262.1) - (AL050090) hypothetical protein [Homo sapiens]		UNCLASSIFIED	264569, 264908, 33109954, 264763, 21908768, 60170394, 18108385, 264563, 264259, 29331828, 35696052, 264909, 265008, 265017, 265018, 18108351, 264288, 21908768, 33657023, 33657109, 264628, 18108374, 35695855, 264634, 264555, 264556, 264557, 264558, 264559
2138	87395446 (4275, 4276)	Novel Protein sim. GBank gi 5174779(emb AAD40696.1) - (U87804) 50 kDa protein [Caulobacter crescentus]		ATPase_associated	264905, 264910, 264591, 55812038, 55811386, 85658542, 264760, 18108351, 18108359, 55811957, 265020, 265021, 33657023, 18108384, 55811576, 83373044, 18108385, 56528486, 264482
2139	84843682 (4277, 4278)	Novel Protein sim. GBank gi 3850821(emb CAA77135) - (Y18350) U2 snRNP auxiliary factor, large subunit [Nicotiana glauca]			

2140	87645655 (4279, 4280)	Novel Protein sim. GBank gij4417293jbjA020418j - (AC007019) unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	264488, 264259, 29331824, 264104, 264109, 264509, 265006, 264759, 265018, 264448, 264288, 21908788, 55811957, 285021, 33657023, 27489265, 35696423, 264636, 264596, 264537, 264559, 264566, 265020, 284893
2141	79623988 (4281, 4282)		UNCLASSIFIED	263978
2142	80041222 (4283, 4284)		UNCLASSIFIED	22278997, 28331827, 264907, 265020, 60432113
2143	84140051 (4285, 4286)	Novel Protein sim. GBank gij2135766jirjS53362 - mucin SAC (clone JER47) - human (fragment)	UNCLASSIFIED	60432113
2144	94320114 (4287, 4288)	Novel Protein sim. GBank gij2078483 (U43200) - antifreeze glycopeptide AFGP polypeptide precursor [Boreogadus saida]	UNCLASSIFIED	65274572, 264259, 28331824, 29331827, 264906, 264908, 284591, 285011, 87188559, 264600, 265019, 284288, 264768, 21906765, 21906767, 55811576, 35698423, 65274791, 22279002
2145	20564305 (4289, 4290)		UNCLASSIFIED	263978
2146	87010515 (4291, 4292)	Novel Protein sim. GBank gij1255871 (U53341) - short region of weak similarity to bovine membrane receptor p63 (PIR-S28503) [Caenorhabditis elegans]	UNCLASSIFIED	264909, 60433356, 264686
2147	80432911 (4293, 4294)	Novel Protein sim. GBank gij3080398jembjCAA18718.1j - (AL022603) putative protein [Arabidopsis thaliana]	UNCLASSIFIED	264907, 264768, 264769, 18108385
2148	80048811 (4295, 4296)	Novel Protein sim. GBank gij728837jispP39194jALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII	UNCLASSIFIED	264593
2149	87362022 (4297, 4298)	Novel Protein sim. GBank gij119863jispP20693jFCE2_MOUSE - LOW AFFINITY IMMUNOGLOBULIN EPSILON FC RECEPTOR (LYMPHOCYTE IGE RECEPTOR) (FC-EPSILON-RII) (CD23)	Contains protein domain (PF00059) - glycoprotein Lectin C-type domain	29331824, 29331826, 35698052, 264758, 87168474, 265018, 52644150, 33657109
2150	94140059 (4299, 4300)	Novel Protein sim. GBank gij542038jembjCAB46679.1j - (AJ243459) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	22278998, 29331822, 29331824, 29331828, 264784, 264769, 21908766, 264486
2151	95353241 (4301, 4302)	Novel Protein sim. GBank gij5689407jdbjBAA82987.1j - (AB028958) KIAA1035 protein [Homo sapiens]		264259, 29331822, 29331824, 29331826, 35698052, 29331828, 264508, 264511, 60433358, 264758, 264596, 33109954, 60174639, 265010, 285011, 87188559, 265017, 265018, 265019, 264448, 264288, 264689, 21906765, 21906766, 21908768, 265020, 60170615, 33657109, 33657182, 33657349, 18108370, 264635, 264557, 60170394, 18108385, 87168518, 22279000
2152	78321840 (4303, 4304)	Novel Protein sim. GBank gij3452473 (AF084205) - serine/threonine protein kinase TAO1 [Rattus norvegicus]	kinase	18108397, 18108398, 265007, 264591, 265011, 18108351, 18108356, 18108374, 18108388

2153	86313371 (4305, 4306)	Novel Protein sim. GBank gll4758704[ref]NP_004216.1[pMASL - MFH-amplified sequences with leucine-rich tandem repeats 1	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	264488, 263994, 52646842, 22278996, 22278998, 22278999, 264259, 29331822, 3569052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 56182435, 264510, 264511, 264512, 264758, 87188474, 87168559, 265017, 265019, 264760, 264288, 264388, 264786, 264687, 264789, 52644229, 21906766, 21906768, 35695917, 33657023, 33657108, 35695855, 264631, 264632, 264635, 264636, 264639, 18108385, 264483, 264584, 264485
2154	87408034 (4307, 4308)	Novel Protein sim. GBank gll225150[prf1209285U - chorion protein B11 [Bombyx mori]		UNCLASSIFIED	56994075, 264094, 265009, 265019, 264288, 21906767, 35695917
2155	87424072 (4309, 4310)			UNCLASSIFIED	18108392, 18108398, 22278996, 264259, 29331824, 265008, 265010, 265011, 265017, 265019, 264288, 264686, 265020, 264693, 264628, 58182323
2156	84285205 (4311, 4312)	Novel Protein sim. GBank gll3970966 (AC004974) - spa-1-like; similar to AF028504 (PID:g2555183) [Homo sapiens]			265007, 264684
2157	87316344 (4313, 4314)	Novel Protein sim. GBank gll1076211[prf1530755 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii]		UNCLASSIFIED	264591
2158	86444218 (4315, 4316)	Novel Protein sim. GBank gll465084[dbj]BAA77027.1 - (AB026190) Kelch motif containing protein [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	dna_rna_bind	29331822, 264112, 265009, 264691, 33657023, 264634
2159	80083729 (4317, 4318)	Novel Protein sim. GBank gll2879825[dbj]BAA248261 - (AB007897) KIAA0437 [Homo sapiens]			264634
2160	16283674 (4319, 4320)				
2161	87739131 (4321, 4322)	Novel Protein sim. GBank gll1504006[dbj]BAA132021 - (D86966) similar to human ZFY protein [Homo sapiens]		UNCLASSIFIED	265008
2162	84319528 (4323, 4324)			UNCLASSIFIED	55274572, 264508, 264805, 264906, 264907, 264908, 52644045, 264809, 265007, 264910, 264591, 264592, 264593, 55812038, 264596, 264758, 265011, 264600, 264762, 264763, 264683, 264784, 264288, 264786, 264686, 264788, 264789, 264689, 265020, 264691, 264628, 264629, 263978, 264632, 264634, 264557, 264838, 264839, 18108385, 264563, 264586, 264587
2163	95417158 (4325, 4326)	Novel Protein sim. GBank gll3876537[emb]CAA982701 - (Z73974) cDNA EST yk29115.3 comes from this gene; cDNA EST yk29115.5 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	56182575, 22278996, 264093, 264683, 33657023, 85274620, 60432113
2164	80569456 (4327, 4328)		Contains protein domain (PF01006) - Hepatitis C virus non-structural protein NS4a	collagen	264603, 264637, 264585

2165	84329189 (4329, 4330)	Novel Protein sim. GBank gij1086794 (U41107) - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	56994075, 22278996, 22278997, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264908, 29331830, 56182435, 265009, 21906754, 33857084, 265011, 265019, 264448, 264288, 264389, 21906765, 21906768, 21906769, 265020, 265021, 264691, 264692, 33657023, 65274620, 35695855, 264556, 60170394, 83373044, 60432113, 22278998, 66714117, 52645156, 22278994, 22278998, 66714117, 29331828, 52644045, 265018, 265019, 264369, 21906765, 21906768, 21906769, 265021, 265022, 264693, 27486282, 35695763, 18108376, 56526486, 87168518, 264567
2166	87618934 (4331, 4332)	Novel Protein sim. GBank gij2706522[emb]CAA75816] - (Y15895) ubiquitin activating enzyme [Drosophila melanogaster]	ubiquitin		56182575, 35695286, 29331824, 29331828, 29146498, 56182435, 265008, 265009, 264592, 264593, 33657402, 33109954, 265011, 265017, 265018, 18108351, 264369, 21906764, 21906765, 21906768, 29148627, 21906769, 52644150, 33657109, 35696423, 18108381, 18108384, 18108385, 60432113, 264567
2167	87716864 (4333, 4334)	Novel Protein sim. GBank gij2224713[dbj]BAA20840] - (AB002384) KIAA0388 [Homo sapiens]		UNCLASSIFIED	56182575, 35695286, 29331824, 29331828, 29146498, 56182435, 265008, 265009, 264592, 264593, 33657402, 33109954, 265011, 265017, 265018, 18108351, 264369, 21906764, 21906765, 21906768, 29148627, 21906769, 52644150, 33657109, 35696423, 18108381, 18108384, 18108385, 60432113, 264567
2168	86999334 (4335, 4336)	Novel Protein sim. GBank gij4321407[gb]AAD15748] - (AF047690) ATP-binding cassette protein M-ABC1 [Homo sapiens]	Contains protein domain (PF00664) - ABC transporter transmembrane region.	transport	66714117, 29331827, 284907, 264511, 264591, 265018, 264764, 264683, 264766, 264768, 264566
2169	87886937 (4337, 4338)				264628, 264555, 264558
2170	94114033 (4339, 4340)	Novel Protein sim. GBank gij5106521[gb]AAD39741.1[AF10536] K-CI cotransporter KCC4 [Homo sapiens]		UNCLASSIFIED	65274572, 56182575, 22278997, 22278998, 264259, 29331825, 264509, 264908, 56182435, 60433438, 55812038, 264598, 55811386, 265019, 264762, 264763, 264448, 264764, 264684, 264288, 264768, 264685, 56181562, 264689, 55811957, 265020, 264535, 264691, 33657109, 60431528, 18108374, 35696423, 55811576, 65274791, 284634, 264639, 264558, 87168518, 60432113, 264584
2171	80194050 (4341, 4342)			UNCLASSIFIED	264369, 265020, 264558
2172	85452460 (4343, 4344)			UNCLASSIFIED	264259, 264558
2173	87036740 (4345, 4346)	Novel Protein sim. GBank gij4309681[gb]AAD15478] - (AC006830) R33423.1 [Homo sapiens]		UNCLASSIFIED	264369
2174	95003288 (4347, 4348)	Novel Protein sim. GBank gij2493778[sp]Q09456[YQ35_CAEEL - PUTATIVE CUTICLE COLLAGEN C09G5.3]			264906, 35695855, 264555, 264557

2175	94325850 (4349, 4350)	Novel Protein sim. GBank gi 1263287 (U47855) - fibronin-3 (Acanus diadematus)		UNCLASSIFIED	264488, 35696286, 20281099, 28331826, 60432289, 35696052, 264109, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265006, 265007, 264512, 265008, 264910, 264581, 264594, 264595, 264596, 264758, 55812038, 265011, 264800, 264603, 264760, 264762, 264448, 264764, 264288, 264766, 264688, 264687, 21906768, 55811957, 35695917, 265020, 265022, 264691, 264892, 33637023, 264693, 264628, 264629, 55811576, 35696423, 65274791, 35695855, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 264639, 264558, 18108385, 60432113, 264563, 264564, 264565, 264566, 264486, 264567
2176	88223392 (4351, 4352)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - III: ALU SUBFAMILY SQ WARNING ENTRY III:	Contains protein domain (PF00805) - Pentapeptide repeats (8 copies)	oncogene	52844507, 52846842, 22278994, 35696286, 22278998, 22278999, 29331828, 29331827, 35696052, 29331828, 33658970, 29331830, 264910, 33657402, 264758, 52844296, 87168559, 265018, 264688, 21906765, 21906767, 21906769, 35695917, 52844180, 264690, 33657023, 33657109, 52845129, 33657182, 27486261, 27486262, 33657349, 18108376, 18108377, 35695855, 87168518, 80432113, 264404, 22279000, 264489
2177	84128942 (4353, 4354)	Novel Protein sim. GBank gi 5454072 ref NP_008416.1 pSLU7 - step II splicing factor SLU7		kinase	18108302, 22278997, 22278999, 264093, 33657402, 265019, 264448, 264766, 264689, 21906767, 21906768, 21906769, 265021, 33657023, 18108370, 18108374, 60432113, 22279002
2178	87601557 (4355, 4356)	Novel Protein sim. GBank gi 473407 (U08215) - NST-1 (Mus musculus)	Contains protein domain (PF00012) - Hsp70 protein	eph	264488, 22278998, 22278999, 29331824, 29331825, 29331826, 29331827, 52844286, 87168474, 18108370, 35695855, 22279002
2178	87318275 (4357, 4358)			UNCLASSIFIED	60424269, 264760, 264628, 264632

2180	95351397 (4358, 4360)	Novel Protein sim. GBank gi 3122317 sp P50648 KMH_B_DICDI - MYOSIN HEAVY CHAIN KINASE B (MHCK B)	Contains protein domain (PF00400) - WD domain, G-beta repeat.	Kinase	52644507, 22278994, 35696286, 22278997, 22278998, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331828, 33656970, 264508, 264509, 264906, 264908, 29331830, 264909, 264510, 265006, 265007, 33657402, 59812038, 21908794, 87168474, 87168559, 265017, 265018, 265019, 264763, 264682, 264683, 264684, 264288, 264686, 21908765, 21908766, 21908769, 265020, 265021, 265022, 52644150, 33657023, 33657109, 27486265, 33657349, 18108374, 35696423, 35695855, 263981, 60170394, 18108385, 56526486, 87168518, 60432113, 22278000, 264482, 264568, 264567, 264486, 29331827, 264369, 18108376, 264584
2181	85764930 (4361, 4362)	Novel Protein sim. GBank gi 3024689 sp Q15942 T2D4_HUMAN - TRANSCRIPTION INITIATION FACTOR TFIID 100 KD SUBUNIT (TAFII-100) (TAFII100)		Kinase	
2182	87637731 (4363, 4364)	Novel Protein sim. GBank gi 5420387 emb CAB46679.1  - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	22278996, 22278997, 22278999, 264259, 29331822, 56182435, 264112, 264764, 264288, 21906767, 21908768, 21908769, 33657109, 18108376, 60170394, 22279000, 22278002
2183	85460649 (4365, 4366)	Novel Protein sim. GBank gi 3873408 gb AAC77482.1  - (U17129) unknown [Rhodococcus erythropolis]			264760
2184	87760690 (4367, 4368)	Novel Protein sim. GBank gi 3114713 (AF081346) - Edp1 protein [Mus musculus]		tnf	29331822, 29331825, 29331826, 56182435, 265011, 264685, 264686, 21906768, 18108370, 264629, 264631, 264636, 264557
2185	87828463 (4369, 4370)	Novel Protein sim. GBank gi 5106958 gb AAD39905.1 AF11361 - (AF113615) FHL1/FH2 domain-containing protein FHOS [Homo sapiens]			29331824, 264907, 66712502, 264757, 265019, 264268, 264692, 56526486
2186	87739227 (4371, 4372)	Novel Protein sim. GBank gi 2864625 emb CAA16972  - (AL021811) putative protein [Arabidopsis thaliana]		ATPase_associated	264259, 29331822, 29331824, 29331826, 56182435, 264592, 55812038, 264760, 264768, 59811957, 33657023, 55811576, 56182323, 264563
2187	87388173 (4373, 4374)			UNCLASSIFIED	35696032, 264905, 264906, 264907, 264908, 264510, 264511, 265008, 264910, 264758, 265018, 264762, 264681, 264766, 264769, 35695917, 264692, 35686423, 264631, 264635, 264637, 18108388, 264588, 264488
2188	87771708 (4375, 4376)	Novel Protein sim. GBank gi 5107818 gb AAD40129.1 AF14941 - (AF149413) contains similarity to histone deacetylases; Plant PF00850. Score=13.3, E=9e-10, N=1 [Arabidopsis thaliana]		histone	18108398, 56994075, 264259, 29331824, 29331825, 66714117, 29331827, 264906, 29331830, 265018, 265020, 265021, 56182323, 264559, 22279000, 22279002
2189	85693573 (4377, 4378)	Novel Protein sim. GBank gi 3452357 (AF075724) - unknown [Legionella pneumophila]	Contains protein domain (PF01596) - O-methyltransferase		22278996, 264259, 29331826, 21906754, 264369, 264288, 263997



2190	87639197 (4379, 4380)	Novel Protein sim. GBank gij132575[sp]P29315[RIN]_RAT - RIBONUCLEASE INHIBITOR		nucleasenhb	22278996, 22278999, 29331822, 29331824, 29331826, 265008, 264910, 60170831, 55912038, 52644296, 265010, 265018, 264685, 264688, 56181562, 21908769, 35695917, 265022, 60170394, 22279000, 29331825, 29331826, 29331830, 264510, 264511, 264910, 264593, 264594, 264596, 264559, 264558
2191	95198928 (4381, 4382)	Novel Protein sim. GBank gij5327002[emb]CAB46272.1] - (Y18503) XAP-5-like protein [Homo sapiens]			
2192	11126310 (4383, 4384)	Novel Protein sim. GBank gij462600[sp]P34400[M110, CAEEL - MIG-10 PROTEIN PH domain	Contains protein domain (PF00169) - PH domain		
2193	94140073 (4385, 4386)	Novel Protein sim. GBank gij5420369[emb]CAB46680.1] - (AJ243460) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	56181686, 29331825, 29331827, 264508, 264909, 265008, 264592, 60432228, 264288, 264684, 264766, 35695917, 33657023, 60431602, 60431528, 55910764, 55811576, 85274791, 35695855, 60431650, 56182323, 60432113, 264592
2194	21418714 (4387, 4388)	Novel Protein sim. GBank gij2773341 (AF040954) - putative protein phosphatase 1 nuclear targeting subunit [Rattus norvegicus]			
2195	88083023 (4389, 4390)	Novel Protein sim. GBank gij2832763[emb]CAA15585.1] - (AL009191) /prediction=(method:: /prediction=(method:: /match=(desc:: /molif=(desc:: [Drosophila melanogaster])		UNCLASSIFIED	22278996, 22278999, 35696052, 265008, 21908764, 265017, 35695917, 265021, 285022, 35695855
2196	95091631 (4391, 4392)	Novel Protein sim. GBank gij5262487[emb]CAB45699.1] - (AL080076) hypothetical protein [Homo sapiens]		collagen	56182575, 35696286, 22278997, 22278999, 264259, 29331822, 66714117, 60432289, 29331827, 35696052, 29331828, 264508, 52644045, 56182435, 264510, 265007, 265008, 265009, 60433438, 55812038, 285010, 265011, 264448, 264288, 264686, 284687, 52644229, 21906765, 21906766, 21908767, 35695917, 265022, 264691, 33657023, 284693, 18108370, 18108376, 35696423, 55811576, 65274791, 35695855, 264636, 56182323, 18108385
2197	95073813 (4393, 4394)	Novel Protein sim. GBank gij4928567[gb]AAD34044.1 AF15180 - (AF151807) CGI-49 protein [Homo sapiens]			284768, 264769, 21906765, 21906766, 21908767, 29148627, 55811957, 35696286, 265020, 22278998, 265021, 264259, 33657023, 284693, 29331824, 35696052, 29331828, 18108370, 35695855, 264113, 265008, 264910, 60432229, 56182323, 33657402, 284758, 83373044, 21908754, 265018, 265019, 22279002, 264482, 264448, 264565, 264288, 264389
2198	88060914 (4395, 4396)	Novel Protein sim. GBank gij3548787 (AC005622) - R39953.1 [Homo sapiens]		UNCLASSIFIED	

2199	85054355 (4397, 4398)	Novel Protein sim. GBank gij2739372 (AC002505) - hypothetical protein [Arabidopsis thaliana]			264105, 264110, 264112, 264688, 55811957, 33657023, 264692, 265967, 20281071, 56526486
2200	87405385 (4399, 4400)	Novel Protein sim. GBank gij3043634[idjBAA2548]-[AB011127] KIAA0555 protein [Homo sapiens]	struct		29331824, 264763, 264768
2201	94316872 (4401, 4402)	Novel Protein sim. GBank gij3913470[idjQ57314]DHBX_ANAPL - PUTATIVE STEROID DEHYDROGENASE SPM2	Contains protein domain (PF00106) - short chain dehydrogenase	dehydrogenase	29331824, 35696052, 264905, 264907, 33657402, 55811386, 265017, 265018, 265019, 264288, 21906768, 35695917, 265020, 265022, 33657023, 33657109, 27486261, 18108370, 35696423, 35695855, 264555, 264556, 83373044, 87168518, 60432113
2202	91672385 (4403, 4404)	Novel Protein sim. GBank gij5262665[emb CAB45767.1]-[AL080166] hypothetical protein [Homo sapiens]	UNCLASSIFIED		264489, 264258, 29331824, 60432289, 35696052, 264905, 264909, 264592, 265017, 265018, 265019, 18108351, 264762, 264440, 264369, 264288, 264766, 21906765, 21906768, 264890, 264891, 264692, 33657109, 264634, 264636, 264555, 264639, 264558, 264559, 83373044, 18108385, 264404, 22279002, 264482
2203	87761832 (4405, 4406)	Novel Protein sim. GBank gij1172845[idjP46629]RB25_RABIT - RAS-RELATED PROTEIN RAB-25	Contains protein domain (PF00071) - Ras family	glycoprotein	52646365, 56994075, 264259, 29331822, 29331826, 29331827, 29331828, 264910, 265010, 265011, 87168559, 265018, 265019, 264605, 264288, 21906769, 35695917, 33657023, 264692, 33657109, 35695763, 18108376, 264638, 22279000, 264556, 264557
2204	88088671 (4407, 4408)	Novel Protein sim. GBank gij121038[idjP29349]GBT3_RAT GUANINE NUCLEOTIDE-BINDING PROTEIN G(T), ALPHA-3 SUBUNIT (GUSTOUCIN ALPHA-3 CHAIN)	Contains protein domain (PF00503) - G-protein alpha subunit	UNCLASSIFIED	
2205	94147589 (4409, 4410)	Novel Protein sim. GBank gij4589480[idjBAA76768.1]-[AB023141] KIAA0924 protein [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_rna_bind	18108394, 18108397, 56182575, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 264806, 265007, 265008, 265009, 80432229, 265010, 265011, 265018, 264683, 264288, 264369, 264698, 21906766, 21906768, 21906769, 264690, 264691, 264693, 18108388, 55811576, 65274791, 264634, 18108381, 18108384, 60432113, 22279002, 264553, 264556, 264591
2206	20620008 (4411, 4412)	Novel Protein sim. GBank gij4557753[idjNP_000372.1]pMID1 - midline 1 protein	Contains protein domain (PF00622) - SPRY domain	UNCLASSIFIED	29331822, 56182181, 29331827, 35696052, 52644045, 265006, 265019, 56181552, 55811957, 265021, 33657023, 35695763, 35695855, 60170394, 60432113, 264568
2207	87787970 (4413, 4414)				
2208	86100830 (4415, 4416)	Novel Protein sim. GBank gij3986746 (AF105228) - tuftelin [Bos taurus]		struct	264906, 265019, 18108351, 21906769, 264112, 265009, 264691, 18108365, 18108374, 264634, 20281166
2209	87800420 (4417, 4418)				

2210	57152407 (4419, 4420)	Novel Protein sim. GBank gij728837/sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		kinase	264603
2211	87341720 (4421, 4422)	Novel Protein sim. GBank gij728837/sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		oncogene	264685, 264686, 18108365, 22279002, 264482
2212	91223924 (4423, 4424)	Novel Protein sim. GBank gij3776027 emb CAA09214  - [AJ010475] RNA helicase [Arabidopsis thaliana]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	22278995, 22278997, 22278999, 264092, 264094, 29331822, 66714117, 29331828, 29331828, 264907, 52644045, 265009, 60170831, 21906754, 87168559, 265017, 265019, 18108351, 264683, 18108354, 264389, 264766, 264687, 52644229, 21906765, 21906766, 21906767, 21906768, 265021, 33657108, 18108370, 18108374, 264636, 56182323, 18108384, 18108387, 87168518, 264565
2213	91219309 (4425, 4426)	Novel Protein sim. GBank gij5420387 emb CAB46678.1  - [AJ243459] proteophosphoglycan [Leishmania major]			56182575, 22278988, 22278997, 35686052, 264905, 66712502, 264908, 264828, 56182435, 264112, 265008, 60431735, 60433438, 21906754, 265010, 265011, 265017, 265018, 265019, 18108351, 264765, 21906765, 21906768, 21906769, 265020, 265021, 264683, 264629, 263974, 263976, 18108379, 55811576, 264556, 264637, 264558, 83373044, 22279002, 264482, 264483

2214	85361453 (4427, 4428)	Novel Protein sim. GBank gjl4504325[re]NP_000173.1pHADH - hydroxyacyl- Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha su	Contains protein domain (PF00725) - 3-hydroxyacyl-CoA dehydrogenase	264488, 52644507, 18108394, 56182575, 22278894, 22278895, 35696288, 56994075, 22278897, 22278898, 22278899, 264490, 60432049, 264258, 52645080, 28331822, 29147620, 29331824, 66714117, 29331825, 60432289, 29331826, 28331827, 35996052, 29331828, 20281100, 264509, 264907, 68712502, 264908, 29331830, 52644045, 56182435, 264510, 265008, 264511, 264512, 265007, 265008, 265009, 60170831, 284583, 60433356, 60433438, 33109954, 33657084, 52644296, 87168474, 265010, 265011, 87168559, 264601, 265017, 265018, 265019, 18108351, 264448, 264682, 264763, 264208, 264687, 52644229, 264689, 21905765, 21906766, 21906767, 21906768, 21906769, 35895917, 265020, 265021, 265022, 264532, 60170815, 264690, 52644150, 264691, 33657023, 264692, 18108384, 33857109, 33857182, 27488282, 27488264, 27488265, 35895763, 18108370, 264629, 60431528, 18108374, 18108376, 55810764, 35696423, 35695855, 264634, 264636, 52644332, 264638, 264558, 60170384, 18108381, 56182323, 83373044, 18108385, 18108387, 18108388, 56528486, 87168518, 60432113, 22279002, 264482, 264584, 264585, 264586, 264509, 265006, 264535, 264538, 87168518
2215	85418206 (4429, 4430)	Novel Protein sim. GBank gjl1947160 (AF000288) - weak similarity to collagens; glycine- and proline-rich [Caenorhabditis elegans]		
2216	87614048 (4431, 4432)	Novel Protein sim. GBank gjl1572802 (U70854) - similar to Enterococcus faecalis TRAB (GI:388288) [Caenorhabditis elegans]	UNCLASSIFIED	264693
2217	80589404 (4433, 4434)	Novel Protein sim. GBank gjl5031707[re]NP_005503.1pGARP - glycoprotein A repetitions predominant	Contains protein domain (PF00560) - Leucine Rich Repeat	264288, 33657109, 264556
2218	85518254 (4435, 4436)	Novel Protein sim. GBank gjl387863[emb]CAA68953] - (Z49128) similar to cAMP-dependant protein kinase; cDNA EST EMBL: T00719 comes from this gene; cDNA EST yk465d8.3 comes from this gene; cDNA EST yk465d8.5 comes from this gene; cDNA EST yk49214.3 comes from this gene; cDNA EST y...	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	35696423, 264563
2219	87614048 (4437, 4438)	Novel Protein sim. GBank gjl1572802 (U70854) - similar to Enterococcus faecalis TRAB (GI:388288) [Caenorhabditis elegans]	Contains protein domain (PF01963) - TraB family	264682, 264683, 264688, 264689, 264693, 18108370, 18108376

2220	85354165 (4439, 4440)	Novel Protein sim. GBank gil4507261[refNP_003145.1]pSTAT - stathin				264488, 18108394, 18108395, 35595286, 264259, 284097, 60432289, 264508, 264905, 284908, 284907, 29331830, 264908, 264909, 264510, 264511, 265007, 264512, 284910, 265008, 264593, 264594, 60433356, 264595, 55812038, 264758, 85658542, 265010, 264601, 264603, 265019, 264605, 264760, 264762, 264448, 264764, 264389, 264766, 18108357, 264768, 264687, 18108358, 264769, 55811957, 264690, 264691, 33657023, 264692, 18108382, 18108388, 264628, 264629, 18108374, 263978, 264634, 264635, 264636, 264637, 264638, 18108385, 264683, 264566, 264486, 264567
2221	88050927 (4441, 4442)	Novel Protein sim. GBank gil3549154 (AC005625) - R27328_1 [Homo sapiens]				
2222	84425892 (4443, 4444)				UNCLASSIFIED	264908, 265020, 35595855
2223	95091849 (4445, 4446)				UNCLASSIFIED	265010, 264685, 264890, 264693, 264628, 263974, 263978, 55811576, 264555, 264636, 83373044, 264483
2224	87388515 (4447, 4448)	Novel Protein sim. GBank gil3876005[emb]CAA84799] - (Z35719) cDNA EST EMBL:D67419 comes from this gene; cDNA EST EMBL:C13853 comes from this gene; cDNA EST EMBL:C11578 comes from this gene; cDNA EST yk234a7.3 comes from this gene; cDNA EST yk234a7.5 comes from this gene; cDNA ES.	Contains protein domain (PF01958) - Domain of unknown function		UNCLASSIFIED	264259, 264508, 56182435, 265006, 265008, 265009, 264757, 21906754, 18108351, 264693, 18108374, 18108385
2225	85749484 (4449, 4450)	Novel Protein sim. GBank gil1255847 (U53338) - C05E11.1 gene product [Caenorhabditis elegans]			transport	22278994, 22278995, 22278999, 52644045, 264600, 265019, 21908785, 21906769
2226	86978953 (4451, 4452)	Novel Protein sim. GBank gil4826524[emb]CAB42852.1] - (ALC49848) hypothetical protein [Homo sapiens]				264259, 29331822, 29331824, 29331825, 28331827, 264508, 284906, 265007, 264691, 264634, 264486
2227	87721135 (4453, 4454)				UNCLASSIFIED	22278999, 265006, 265008, 18108354, 29148629, 29148784, 27486281, 18108374, 264637, 18108384
2228	91227337 (4455, 4456)	Novel Protein sim. GBank gil606976 (U16800) - ribonucleoprotein [Xenopus laevis]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)		dna_rna_bind	264091, 264092, 264094, 29331822, 29331825, 66714117, 264693, 263972, 264639, 83373044, 264563
2229	88060931 (4457, 4458)	Novel Protein sim. GBank gil3549155 (AC005625) - R27328_2 [Homo sapiens]			UNCLASSIFIED	

2230	95342915 (4459, 4460)	Novel Protein sim. GBank gij226154[prf]11412350A - DNA polymerase [Human adenovirus type 2]	UNCLASSIFIED	264489, 264768, 52644507, 264769, 21908765, 21906766, 21906767, 21906769, 22278995, 35695917, 22278996, 22278997, 22278998, 22278999, 265021, 264259, 52645129, 29331827, 264508, 264509, 264907, 18108370, 18108374, 35696423, 35695855, 265007, 264910, 264555, 33657402, 21906754, 18108387, 265010, 265018, 265019, 264760, 264288, 264587, 264563
2231	88060937 (4461, 4462)	Novel Protein sim. GBank gij3549154 (AC005625) - R27328 - 1 [Homo sapiens]	UNCLASSIFIED	
2232	87762581 (4463, 4464)	Novel Protein sim. GBank gij5281316[gb]AAD41476.1[AF133124] transcription factor IIC83 [Homo sapiens]	transcript factor	18108394, 56182575, 22278995, 35696286, 22278997, 22278999, 264259, 29331827, 35696052, 264907, 56182435, 265008, 265007, 265008, 264910, 264758, 59812038, 264603, 265018, 265019, 18108351, 284882, 264764, 264683, 264369, 264288, 264686, 264887, 264688, 21906765, 21906766, 21906767, 21906769, 29148629, 35695917, 264690, 52644150, 264691, 33657023, 264693, 18108370, 18108374, 55811576, 35695855, 264639, 18108385, 264564, 264908, 33657402, 265018, 264288, 264686, 265020, 264635, 18108395
2233	87755292 (4465, 4466)	Novel Protein sim. GBank gij4249733[gb]AAD13780 - (AF109377) [Mus musculus]		56182435, 264369, 284888, 21906765, 265020, 264693, 264556, 58526488
2234	87771817 (4467, 4468)	Novel Protein sim. GBank gij1706559[sp]P54332[EAS_DROME - ETHANOLAMINE KINASE (EASILY SHOCKED PROTEIN)]	kinase	22278997, 264563
2235	91012316 (4469, 4470)	Novel Protein sim. GBank gij4972734[gb]AAD34762.1 - (AF132174) unknown [Drosophila melanogaster]	UNCLASSIFIED	264589, 264687, 264769, 265022, 264259, 60432049, 284891, 29331826, 60432289, 20281149, 284906, 264907, 264511, 265008, 265009, 264634, 284635, 284636, 264555, 264556, 264557, 264558, 60433356, 264595, 264559, 60433438, 60432113, 264761, 264762, 264763, 264764
2236	88003131 (4471, 4472)	Novel Protein sim. GBank gij1082675[prf]B53814 - p20 protein - human	Contains protein domain (PF00400) - WD domain, G-beta repeat Contains protein domain (PF00011) - eph Hsp20/alpha crystallin family	264488, 264769, 21906765, 21906766, 21908767, 21906768, 21906769, 22278995, 22278996, 22278997, 22278998, 265020, 265021, 22278999, 264259, 29331824, 29331826, 29331827, 35695763, 18108376, 35695855, 265007, 60432229, 33657402, 60433356, 60433438, 83373044, 18108385, 21906754, 18108387, 60432113, 22279000, 265019, 22279002, 264482, 18108351, 264288
2237	91012318 (4473, 4474)	Novel Protein sim. GBank gij4972734[gb]AAD34762.1 - (AF132174) unknown [Drosophila melanogaster]	kinase	

2238	94998857 (4475, 4478)		Contains protein domain (PF00288) - Viral coat protein	264509, 264807, 264829, 264634, 264564
2239	87798688 (4477, 4478)			28331825, 265009, 264369, 33657109, 18108370, 18108374, 264557, 264559
2240	94121471 (4478, 4480)	Novel Protein sim. GBank gll2882311 (AF051240) - probable ubiquitin-conjugating enzyme E2 [Picea mariana]	Contains protein domain (PF00179) - Ubiquitin-conjugating enzyme	264488, 65274572, 56182375, 35608286, 22278997, 22278999, 264259, 29331827, 35896052, 264508, 52844045, 56182435, 284511, 265007, 265008, 265009, 60433356, 60433438, 55812038, 21906754, 33657084, 55811388, 265018, 265019, 18108351, 264683, 264288, 264788, 264687, 264689, 264769, 21906765, 21908788, 21906769, 35895917, 265021, 265022, 80170615, 52644150, 33657023, 33657182, 33657349, 35895763, 18108370, 35896423, 35895855, 87168518, 22278000
2241	80091951 (4481, 4482)		UNCLASSIFIED	284693, 264629
2242	91228075 (4483, 4484)	Novel Protein sim. GBank gll2494312 [sp70541] E2BG_RAT - TRANSLATION INITIATION FACTOR EIF-2B GAMMA SUBUNIT (EIF-2B GDP-GTP EXCHANGE FACTOR)	synthase	22278995, 22278996, 22278997, 22278998, 264259, 29331822, 29331824, 29331826, 29331827, 29331828, 264509, 265007, 265009, 264598, 21908754, 265010, 265011, 265017, 265018, 265019, 264448, 264389, 264288, 52844229, 21906765, 21908788, 21908787, 21908768, 21906769, 265020, 265021, 33657109, 27486282, 27486284, 18108374, 35895855, 264634, 264637, 56182323, 83373044, 56526486, 87168518, 264564
2243	78902026 (4485, 4486)		UNCLASSIFIED	285008
2244	85723527 (4487, 4488)	Novel Protein sim. GBank gll2291143 (AF016417) - Similar to BZIP transcription factor [Caenorhabditis elegans]	UNCLASSIFIED	264604
2245	95318545 (4489, 4490)	Novel Protein sim. GBank gll470340 (U00043) - similar to beta-mannosyltransferase [Caenorhabditis elegans]	Contains protein domain (PF00534) - Glycosyl transferases group 1	52845156, 22278995, 22278996, 22278997, 22278999, 29331822, 29331824, 29331827, 264907, 264512, 60433438, 264758, 21908754, 265011, 264603, 264764, 264687, 21908787, 21906788, 21906769, 55811957, 265022, 264691, 264629, 35896423, 264638, 18108387, 60432113, 22279000, 22278002, 264566

2248	94848710 (4491, 4492)	Novel Protein sim. GBank gij496096[dbj]BAA78326.1] - (AB028058) activator of S phase Kinase [Homo sapiens]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	65274572, 22278995, 35696286, 22278996, 22278997, 22278998, 284259, 35696052, 264106, 284905, 284907, 285006, 285007, 265008, 60433438, 33109854, 87168559, 265018, 265019, 264288, 21906765, 21906767, 21906768, 21906769, 55811957, 35695917, 285020, 285022, 27486264, 18108370, 18108374, 65274791, 35695655, 60432113
2247	87862542 (4493, 4494)	Novel Protein sim. GBank gij854065[emb]CAA58337] - (X83413) U88 [Human herpesvirus 6]	UNCLASSIFIED		52645156, 52646365, 52645080, 35696052, 33655970, 52646317, 33657084, 265017, 21906768, 21906769, 35695917, 33657109, 52645129, 33657182, 27486261, 27486262, 33657349, 27486265, 18108387
2246	95412996 (4495, 4496)	Novel Protein sim. GBank gij4758502[ref]NP_004123.1]pHABP - hyaluronan-binding protein 2	Contains protein domain (PF00089) - Trypsin	calhepsin	264488, 264259, 264907, 29331830, 264909, 265007, 265009, 264595, 21906754, 65274444, 264603, 265019, 264762, 264448, 264288, 284689, 21906766, 55811957, 265021, 284691, 18108374, 264634, 264635, 264636, 264555, 264638, 264557, 264558, 264559, 18108383, 83373044, 18108385, 264486
2249	94685662 (4497, 4498)	Novel Protein sim. GBank gij4038461 (AF107772) - TcSt11 [Trypanosoma cruzi]	Contains protein domain (PF00515) - TPR Domain	eph	264786, 264628, 264636, 264637
2250	79827508 (4499, 4500)	Novel Protein sim. GBank gij3738140[emb]CAA21241] - (AL031852) val-Htna synthetase, mitochondrial precursor [Schizosaccharomyces pombe]	UNCLASSIFIED		264908, 18108374
2251	87385863 (4501, 4502)	Novel Protein sim. GBank gij3218467[emb]CAA07090.1] - (AJ006528) putative phosphatase [Gallus gallus]	UNCLASSIFIED		284259, 35696052, 264508, 56182435, 265009, 264592, 264593, 264780, 264448, 264684, 264288, 264690, 264628, 55811976, 264555, 264556, 264557, 264558, 264559, 264566
2252	87735867 (4503, 4504)	Novel Protein sim. GBank gij4929325[gb]AAD33953.1]AF14531 - (AF145316) vacuolar proton pump delta polypeptide [Homo sapiens]	Contains protein domain (PF01813) - ATP synthase subunit D	synthase	264092, 264094, 264259, 28331822, 66714117, 29331828, 264102, 264103, 264104, 264105, 264109, 264112, 264511, 265007, 60433356, 265010, 18108351, 21906767, 21906768, 264691, 263974, 263977, 264486, 264567
2253	91010703 (4505, 4506)		UNCLASSIFIED		65274572, 265019



2254	95320031 (4507, 4508)	Novel Protein sim. GBank gi 4502847 ref NP_001271.1 pCIRB - cold inducible RNA- binding protein	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	dna_ma_bind	264569, 18108394, 18108398, 56182575, 56894075, 35696286, 22278999, 264094, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 60432288, 29331827, 29331828, 35696032, 264109, 264508, 264509, 264906, 264907, 29331830, 66712502, 264908, 264909, 264510, 265008, 264511, 265007, 265008, 265009, 60170931, 60432228, 60433396, 60433438, 264758, 85658542, 265010, 265011, 87168559, 265017, 265018, 265019, 264448, 264764, 264288, 264369, 264768, 264686, 264768, 264769, 21906765, 21906767, 55811937, 264691, 33657023, 264692, 18108362, 85274620, 263969, 264928, 18108370, 60431528, 263972, 264629, 18108372, 18108377, 18108379, 55811576, 35696423, 35695855, 264630, 264634, 264635, 264636, 264556, 263981, 264638, 56182323, 60170394, 264558, 18108381, 18108382, 83373044, 18108385, 87168518, 60432173, 22279002, 264482, 264564, 264565, 264486, 264567, 18108391
2255	91010546 (4509, 4510)	Novel Protein sim. GBank gi 5541865 emb CAB51072.1  - (AL096858) hypothetical protein [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	stud1	65274572, 56182575, 22278997, 22278999, 264259, 29331822, 29331825, 29331826, 29331827, 29331828, 264508, 264905, 264906, 264907, 66712502, 264908, 56182435, 264510, 264511, 265008, 264593, 264595, 21906754, 33109954, 87168474, 265011, 265017, 265019, 264682, 264764, 264369, 264288, 264768, 264685, 264686, 264768, 21906765, 21906766, 21906768, 21906769, 265020, 60170615, 52644150, 264690, 264692, 264693, 33657109, 33657349, 264632, 264638, 52644332, 56182323, 22279000, 22279002, 264768, 264689, 18108374
2256	87020531 (4511, 4512)	Novel Protein sim. GBank gi 3327174 dbj BAA31655  - (AB014580) KIAA0680 protein [Homo sapiens]		UNCLASSIFIED	
2257	80088235 (4513, 4514)			UNCLASSIFIED	22278996, 22278999, 264681, 21906765, 21906766, 264567
2258	88090516 (4515, 4516)	Novel Protein sim. GBank gi 3025446 (AC004528) - R32184_2 [Homo sapiens]	Contains protein domain (PF00060) - Ligand-gated ion channel	misc_channel	264908, 264592, 264764

2259	95364155 (4517, 4518)	Novel Protein sim. GBank gij4884140[emb]CAB43276.1] - (AL050110) hypothetical protein [Homo sapiens]		UNCLASSIFIED	18108396, 65274572, 56182575, 22278997, 22278998, 264269, 28331822, 29331827, 284805, 68712502, 284908, 264909, 58182435, 285007, 265008, 60432229, 33657084, 87168559, 18108351, 264448, 264683, 264288, 264389, 58181582, 265021, 60170615, 264690, 33657109, 60431528, 18108374, 52644332, 56182323, 18108385, 22279000, 22279002, 264482
2260	88084119 (4519, 4520)	Novel Protein sim. GBank gij3080863 (AC004614) - similar to f-spondin proteins AB008086 (P1D:2529225) [Homo sapiens]	Contains protein domain (PF00090) - Thrombospondin type 1 domain	oxidase	58182575, 265020, 264905, 264906, 264908, 35696423, 264511, 264635, 55812038, 264758, 265018, 265019, 264605, 264760, 284583
2261	88074157 (4521, 4522)	Novel Protein sim. GBank gij3334526[emb]CAA18138] - (AL021306) predicted using FGENEH [Homo sapiens]		UNCLASSIFIED	
2262	91639292 (4523, 4524)	Novel Protein sim. GBank gij487759g[AD31421.1]AF12444 - (AF124440) MAGE tumor antigen D1 [Homo sapiens]			56182575, 22278998, 28331822, 28331825, 60432288, 28331827, 35696052, 264508, 68712502, 52644045, 56182435, 265008, 265009, 60433356, 55812038, 265010, 265017, 265019, 264288, 264369, 21906765, 21906767, 55811957, 35695917, 52644150, 33657023, 33657109, 55811576, 65274791, 56182323
2263	87802495 (4525, 4526)	Novel Protein sim. GBank gij3341697 (AC003872) - hypothetical protein [Arabidopsis thaliana]			22278994, 22278997, 264907, 264828, 52644150, 18108381, 264893, 18108374
2264	87756525 (4527, 4528)	Novel Protein sim. GBank gij1657601 (U66220) - unknown [Nannocystis exedens]		UNCLASSIFIED	264686, 284488, 284768, 284769, 264691, 264508, 284905, 264509, 264908, 264907, 264908, 264909, 35695855, 284510, 264511, 264512, 265007, 265008, 264838, 264839, 264757, 264758, 18108385, 265011, 264760, 264584, 264585, 264764, 264586, 264486, 264766
2265	86918563 (4529, 4530)	Novel Protein sim. GBank gij477072[pri]A48018 - mucin 7 precursor, salivary - human	Contains protein domain (PF00096) - Zinc finger, C2H2 type	UNCLASSIFIED	264689, 264910, 264764
2266	87773458 (4531, 4532)	Novel Protein sim. GBank gij3150479 (AF067212) - partial CDS [Caenorhabditis elegans]	Contains protein domain (PF01305) - Ribosomal protein L15 amino terminal region	ribosomal prot	22278995, 22278997, 22278999, 284259, 265005, 265007, 265009, 60433438, 21906754, 265010, 265011, 265017, 264448, 264683, 264288, 264689, 21906785, 21906788, 35695917, 265021, 18108374, 284638, 22279000, 22279002, 264566, 284487

2267	87395838 (4533, 4534)	Novel Protein sim. GBank gj13560229[emb/CAA20697.1] - (ALD31530) hypothetical protein [Schizosaccharomyces pombe]		UNCLASSIFIED	35696286, 264259, 28331824, 29331825, 35696052, 29331828, 264905, 284509, 264907, 264908, 264909, 264512, 285009, 284910, 264593, 33657402, 265010, 265018, 264782, 264448, 284288, 264369, 264768, 52644229, 35699917, 264691, 33657023, 18108392, 33657109, 35696423, 284834, 18108381, 87168518, 264566
2268	85693867 (4535, 4536)	Novel Protein sim. GBank gj1728832[spP39189/ALU2_HUMAN - III] ALU SUBFAMILY SB WARNING ENTRY IIII	cadherin		284488, 264259, 284509, 264595, 265010, 285017, 264766, 18108385, 264486
2269	86177977 (4537, 4538)	Novel Protein sim. GBank gj103418[pit/S17885 - Tcd37 protein - fruit fly (Drosophila melanogaster)]	UNCLASSIFIED		56182575, 60432048, 265007, 285009, 264591, 87168559, 264605, 18108351, 21906764, 265020, 264628, 60431528, 264638, 18108385, 18108387, 60432113, 264783
2270	80410327 (4539, 4540)				264609, 56182435, 265008, 55812038
2271	91010392 (4541, 4542)			cyto450	55811957, 33657023, 264693, 33657109, 55810784, 55811576, 56182323
2272	84208220 (4543, 4544)			UNCLASSIFIED	264905, 264908
2273	95014271 (4545, 4546)	Novel Protein sim. GBank gj14176370 (AC005058) - similar to calcium-independent phospholipase A2; similar to AC004392 (PIDg3367519) [Homo sapiens]	Contains protein domain (PF00462) - Glutaredoxin		52645156, 22278986, 22278989, 60432048, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 264909, 285008, 264593, 60433438, 21906784, 265018, 264888, 21906785, 21906786, 21906767, 21906769, 285021, 265022, 60170615, 264691, 33657023, 264693, 33657109, 27486284, 18108376, 35696423, 35696655, 264630, 52644332, 264558, 56182323, 22279002
2274	81640217 (4547, 4548)	Novel Protein sim. GBank gj1480112[emb/CAA67961] - (X99842) HP1-BP74 protein [Mus musculus]	Contains protein domain (PF00538) - linker histone H1 and H5 family	histone	52645156, 22278997, 22278999, 52645080, 29331824, 29331825, 29331826, 29331827, 29331828, 264905, 264908, 52644045, 264511, 285008, 285009, 60170831, 264591, 21906784, 33108854, 285011, 265018, 18108351, 264448, 264288, 264684, 264768, 21906785, 21906786, 21906787, 21906788, 52644150, 264693, 18108384, 35695763, 18108374, 35696423, 264634, 264557, 284638, 52644332, 83373044, 18108385, 58528488, 87168518, 22279002
2275	88022501 (4549, 4550)	Novel Protein sim. GBank gj13165406 (AC004755) - (os37502_2 [Homo sapiens])	Contains protein domain (PF00122) - E1-E2 ATPase	transport	
2276	11287447 (4551, 4552)			UNCLASSIFIED	264555, 264556

2277	86084123 (4553, 4554)	Novel Protein sim. GBank gii2880079 (AC004142) - similar to murine leucine-rich repeat protein; possible role in neural development by protein-protein interactions; 93% similarity to D49802 (PIL:g1368906) [Homo sapiens]	Contains protein domain (PF00560) - glycoprotein Leucine Rich Repeat	22278999, 35686052, 265008, 265019, 264369, 265020, 265022, 55810764, 264404, 22278002
2278	94133078 (4555, 4556)	Novel Protein sim. GBank gii2618702 (AC002510) - unknown protein [Arabidopsis thaliana]		UNCLASSIFIED
2279	80419375 (4557, 4558)	Novel Protein sim. GBank gii119714sp13883EXTN_TOBAC - EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)		UNCLASSIFIED
2280	84239723 (4559, 4560)			
2281	95283048 (4561, 4562)	Novel Protein sim. GBank gii4240289[idj]BAA74928.1] - (AB020712) KIAA0905 protein [Homo sapiens]	Contains protein domain (PF00400) - transport WD domain, G-beta repeat	265008, 33109854, 265010, 265019, 265020, 264092, 264259, 29331822, 29331824, 29331826, 35686052, 264107, 264906, 264909, 52644045, 265008, 33857402, 60433356, 264758, 265011, 265019, 264681, 264683, 264684, 264686, 21908785, 21908787, 21908788, 21908789, 60170615, 264690, 52644150, 18108362, 264892, 18108368, 18108374, 263978, 264831, 18108381, 264559, 18108385, 56528486, 22279000, 264566, 264567
2282	87602829 (4563, 4564)	Novel Protein sim. GBank gii1537070 (U83840) - nucleoporin p54 [Rattus norvegicus]		UNCLASSIFIED
2283	95362366 (4565, 4566)	Novel Protein sim. GBank gii2495728sp1Q82559[Y281_HUMAN - HYPOTHETICAL PROTEIN KIAA0281 (HA6725)]		UNCLASSIFIED

2284	95414955 (4567, 4568)	Novel Protein sim. GBank gij2496797ipq64311PNAD_MOUSE - PROTEIN N- TERMINAL ASPARAGINE AMIDOHYDROLASE (PROTEIN NH2-TERMINAL ASPARAGINE DEAMIDASE) (NTN- AMIDASE) (PNAD) (PROTEIN NH2-TERMINAL ASPARAGINE AMIDOHYDROLASE) (PNAA)			60424179, 52644507, 18108394, 52646842, 22278994, 35696286, 22278996, 22278997, 22278999, 264259, 60432049, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 35696052, 29331830, 52644045, 56182435, 33657402, 60433436, 33109954, 21906754, 85658542, 87168559, 265018, 265019, 55811150, 264682, 264369, 21906765, 21908766, 21908767, 21908768, 21908769, 55811857, 35695917, 285020, 285021, 60170615, 33657023, 33657182, 27486282, 27486284, 27486285, 18108376, 55810764, 35696423, 35695855, 60431850, 87168518, 60432113, 264482, 264584, 35696052, 264905, 264907, 264908, 264909, 264512, 265009, 264910, 264595, 264760, 18108351, 264682, 264763, 264685, 264766, 264688, 264768, 264693, 264629, 35695855, 264631, 264634
2285	87781484 (4569, 4570)	Novel Protein sim. GBank gij3342234 (U83909) - nuclear antigen EBNA-1 [Cercopithecine herpesvirus 15]	collagen		
2286	87737825 (4571, 4572)	Novel Protein sim. GBank gij3873414 (U00043) - similar to D. melanogaster trithorax protein [Caenorhabditis elegans]	kinase		35696286, 56182435, 60170831, 264591, 60432229, 264592, 264593, 264594, 264595, 55812038, 264596, 87168474, 35695917, 264692, 55811576, 264555, 264557, 264682
2287	82986696 (4573, 4574)	Novel Protein sim. GBank gij630905iprjS42731 - collagen alpha 1 chain - sea urchin [Hemicentrotus pulcherrimus] (fragment)	UNCLASSIFIED	Contains protein domain (PF01391) - Collagen triple helix repeat (20 copies)	
2288	84133083 (4575, 4576)	Novel Protein sim. GBank gij728832ispjP39189/ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII	tm7		22278995, 56994075, 22278997, 22278999, 264259, 60432289, 264508, 264512, 265008, 33657402, 265017, 265018, 265019, 18108351, 264448, 21906765, 21906766, 21906767, 21908768, 33857023, 264557, 22279000, 22279002
2289	88084133 (4577, 4578)	Novel Protein sim. GBank gij2887497 (AC004144) - R34001_1 [Homo sapiens]	kinase	Contains protein domain (PF00400) - WD domain, G-beta repeat	
2290	88084137 (4579, 4580)	Novel Protein sim. GBank gij2887497 (AC004144) - R34001_1 [Homo sapiens]	UNCLASSIFIED	Contains protein domain (PF00400) - WD domain, G-beta repeat	264683
2291	84295281 (4581, 4582)	Novel Protein sim. GBank gij3253120 (AC005175) - R31449_3 [Homo sapiens]	strut		18108394, 264907, 265006, 265009, 33109954, 52646317, 265010, 18108351, 264681, 264686, 264692, 18108370, 18108374, 18108385

2282	94328834 (4583, 4584)	Novel Protein sim. GBank gll4803672[emb](CAB42643.1) - (AJ133789) nuclear transport receptor [Homo sapiens]		UNCLASSIFIED	56182575, 35686286, 56894075, 29331824, 29331825, 35686052, 56182435, 60433438, 55812038, 33109954, 87168474, 87168559, 265018, 18108351, 264763, 264448, 264389, 264288, 58181562, 264769, 21906765, 21906766, 21906767, 21906769, 265021, 265022, 33657023, 264693, 65274620, 33657109, 27486264, 264628, 55810764, 55811576, 35695855, 56182323, 56526486, 87168518, 22278000, 264567
2283	87759213 (4585, 4586)	Novel Protein sim. GBank gll3252981 (AF068921) - Ras-binding protein SUR-8 [Mus musculus]	Contains protein domain (PF00560) - Leucine Rich Repeat	struct	264488, 18108397, 35686286, 264082, 264259, 29331822, 29331826, 264906, 264908, 264511, 264512, 265009, 264910, 18108351, 264764, 264369, 264288, 264685, 264766, 265020, 265022, 264534, 35696423, 264631, 264637, 18108381, 56182323, 264639, 18108385, 264404, 264563, 264585
2284	86693580 (4587, 4588)	Novel Protein sim. GBank gll2082680 (U88964) - HEM45 [Homo sapiens]	Contains protein domain (PF00929) - Exonuclease	nuclease	22278997, 22278998, 22278999, 264259, 29331822, 60432289, 29331828, 35696052, 265018, 264684, 264288, 264686, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 264692, 33657109, 18108376, 35696423, 35695855, 264634, 22279000, 22279002, 264563, 264486
2285	94312200 (4589, 4590)			UNCLASSIFIED	22278998, 60432289, 264682, 264683, 264689, 18108374
2286	80030781 (4591, 4592)				263974, 263978
2297	94321251 (4593, 4594)	Novel Protein sim. GBank gll5689501[db](BAA83034.1) - (AB029005) KIAA1082 protein [Homo sapiens]		transcriptfactor	264488, 65274572, 56182575, 22278997, 22278999, 264259, 29331822, 29331824, 29331826, 29331828, 35686052, 264807, 264908, 52844045, 56182435, 264112, 265006, 265007, 264810, 265009, 60433356, 33657402, 264595, 55812038, 21906754, 265011, 265018, 265019, 264448, 264764, 264288, 264766, 21906765, 21906767, 21906768, 21906769, 55811957, 265020, 265021, 264690, 264691, 33657023, 264692, 264693, 33657109, 55811576, 56182323, 50170364, 83373044, 18108385, 56526486, 264564, 264488

2298	95312207 (4595, 4596)	Novel Protein sim. GBank gij3875051[emb]CAB028491 - (Z81030) predicted using GeneIndex; similar to collagen; cDNA EST EMBL D65564 comes from this gene; cDNA EST EMBL D69046 comes from this gene; cDNA EST yk366b12.3 comes from this gene; cDNA EST yk366b12.5 comes from this gene ...	collagen	60424178, 56181686, 22278995, 35686286, 22278996, 22278998, 22278999, 264490, 264259, 29331822, 29331824, 66714117, 60424268, 35686052, 29331828, 68712502, 56182435, 264510, 265008, 60433438, 21908754, 33109954, 55811388, 265010, 265018, 55811150, 264762, 18108351, 264682, 264683, 264288, 264684, 264688, 264688, 56181562, 264689, 21908766, 21908767, 29148629, 55811857, 29148784, 35695917, 265020, 18108362, 33657023, 18108384, 33657109, 60431602, 18108370, 60431528, 18108374, 55810784, 35686423, 35695855, 264630, 264634, 60431850, 18109380, 56182323, 264558, 83373044, 18108385, 60432113, 22278000, 264482, 264567, 264486
2299	80193720 (4597, 4598)	Novel Protein sim. GBank gij2443886 (AC002294) - Unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	264369
2300	94124348 (4599, 4600)			284488, 22278996, 22278998, 264259, 29331824, 66714117, 35686052, 264508, 284805, 264806, 264907, 264808, 264809, 265008, 264910, 265009, 264758, 265010, 87168559, 264600, 265018, 264760, 264762, 18108351, 264764, 264768, 264768, 264769, 21908766, 21908767, 35695917, 265021, 264691, 33657023, 35695763, 18108370, 18108374, 35686423, 35685555, 264831, 264636, 264638, 18108385, 22278002, 264563
2301	91235725 (4601, 4602)	Novel Protein sim. GBank gij2143637[pri]B4-505 - calcium-dependent actin-binding protein - rat	struct	264908, 264758, 265017, 21906765, 83373044, 264563
2302	88084141 (4603, 4604)	Novel Protein sim. GBank gij2887497 (AC004144) - R34001 - 1 [Homo sapiens]	UNCLASSIFIED	52644045, 265019, 264288, 33657023, 18108370, 18108385
2303	94141439 (4605, 4606)	Novel Protein sim. GBank gij4884194[emb]CAB43220.1 - (AL049946) hypothetical protein [Homo sapiens]	Contains protein domain (PF00047) - struct Immunoglobulin domain	264259, 60432049, 264807, 264909, 264910, 60432229, 33657402, 265011, 265018, 264762, 264448, 264769, 264637, 264638, 83373044, 264488
2304	94840434 (4607, 4608)	Novel Protein sim. GBank gij2494162[sp]Q10005[YRY1_CAEEL - HYPOTHETICAL 39.9 KD PROTEIN T15H8.1 IN CHROMOSOME II PRECURSOR	UNCLASSIFIED	264259, 29331824, 21906767, 33657182, 33657349
2305	90935911 (4609, 4810)	Novel Protein sim. GBank gij4972686[gb]AA034738.1 - (AF132150) unknown [Drosophila melanogaster]		85274372, 22278998, 264808, 265006, 21908769, 264891, 264486

2306	95334940 (4611, 4612)	Novel Protein sim. GBank gi 4929595 gb AA034043.1 AF15180 - (AF151805) CGI-48 protein [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat	Kinase receptor	264488, 22278995, 22278998, 35686286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 35689052, 284508, 264805, 284807, 29331830, 264908, 264909, 264510, 265008, 264511, 265007, 264512, 265008, 265009, 33657402, 21905754, 85658542, 265010, 265011, 264601, 265017, 265018, 264604, 265019, 18108351, 264448, 264288, 264766, 264769, 21908765, 21908768, 21908767, 21908768, 29148629, 28148784, 35689517, 265020, 265021, 265022, 33657023, 264682, 18108370, 18108374, 18108376, 35688423, 35688585, 264630, 284634, 284635, 284636, 284637, 264638, 264639, 18108382, 18108385, 18108387, 284563, 284586, 264488, 264828
2307	78415283 (4613, 4614)	Novel Protein sim. GBank		UNCLASSIFIED	264828
2308	87608408 (4615, 4616)	Novel Protein sim. GBank gi 4758732 ref NP_004522.1 pmOCS - molybdenum cofactor synthesis 2		synthase	35686286, 264259, 29331822, 29331824, 264112, 264512, 264757, 21908754, 264288, 264890, 27488284, 284631, 284634, 284404
2309	95357218 (4617, 4618)	Novel Protein sim. GBank gi 3878059 emb CA817070  - (Z99942) cDNA EST EMBL:D73444 comes from this gene; cDNA EST EMBL:D70905 comes from this gene; cDNA EST EMBL:D72208 comes from this gene; cDNA EST EMBL:D75030 comes from this gene; cDNA EST EMBL:D72844 comes from this gene; cDN...		glycoprotein	18108387, 22278998, 22278997, 22278998, 22278999, 60432049, 29331822, 29331826, 60432288, 66712502, 60432229, 60433356, 60433438, 65274444, 265010, 264600, 264681, 264448, 284683, 264288, 21908766, 21908768, 265020, 264691, 284692, 284693, 65274620, 65274791
2310	28601658 (4619, 4620)	Novel Protein sim. GBank gi 2137337 pir j48281 - gene	Contains protein domain (PF00013) - KH domain	UNCLASSIFIED	264508
2311	87721189 (4621, 4622)	mCBP protein - mouse		transcript factor	18108397, 56182575, 22278998, 56994075, 264259, 29331824, 29331827, 264508, 264807, 58182435, 284510, 284511, 285008, 264512, 285007, 285008, 265009, 60433438, 33108854, 265010, 285011, 284603, 285017, 18108351, 284782, 284683, 284288, 264369, 264686, 33657023, 20281149, 20281089, 264628, 263972, 55811578, 35698423, 20281071, 264632, 264636, 18108385, 18108387, 87188518, 22279000, 264583, 264488



2312	87549681 (4623, 4624)	Novel Protein sim. GBank gij2911264 (AC002550) - Unknown gene product [Homo sapiens]			56182575, 56994075, 35696286, 22278996, 22278997, 22278998, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 52844045, 60433356, 55812038, 33109954, 21906754, 87168474, 265018, 18108351, 264288, 52844229, 21906765, 21906766, 21906767, 21906768, 21906769, 265022, 60170615, 33657023, 27488281, 27488284, 35695855, 18108385, 22278900, 22278902, 263981
2313	80042533 (4625, 4626)	Novel Protein sim. GBank gij3043626(dj)BAA25471 - (AB011123) KIAA0551 protein [Homo sapiens]		UNCLASSIFIED	52844507, 52846365, 52846842, 22278996, 22278997, 22278998, 264259, 52645080, 29331822, 29331824, 60432289, 29331828, 29331827, 29331828, 35696052, 33656970, 52644045, 265008, 264593, 60433356, 60433438, 264758, 33109954, 265010, 265017, 265018, 265019, 284288, 264369, 21906765, 21906766, 21906768, 35695917, 52644150, 33657023, 33657109, 52645129, 33657349, 35695763, 18108374, 35696423, 35695855, 52844332, 22278900, 22278902, 264563, 264567
2314	94313401 (4627, 4628)	Novel Protein sim. GBank gij5596714(emb)(CAB51401.1) - (AL035398) dJ796117.2 (CGI-51) [Homo sapiens]			264905, 264906, 264767, 264768, 264693, 55811576, 264635, 66182323, 18108385
2315	80430119 (4629, 4630)			UNCLASSIFIED	52844507, 52645156, 52646365, 22278994, 22278995, 22278996, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 52645080, 29331824, 28331825, 29331826, 29331827, 35696052, 29331828, 264906, 264907, 28331830, 52844045, 56182435, 264511, 265007, 265008, 265009, 60170831, 60433438, 21906754, 52646317, 33109954, 33657084, 52844296, 87168474, 265010, 87168559, 265017, 285018, 285019, 264681, 264763, 264448, 264683, 264369, 52844229, 21906764, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 52844150, 33657023, 18108362, 52845129, 33657182, 33657349, 35695763, 18108370, 18108376, 35696423, 35695855, 264531, 264558, 52844332, 83373044, 18108385, 18108387, 87168518, 60432113, 22278900, 264568, 264567
2316	94312191 (4631, 4632)	Novel Protein sim. GBank gij5531827(gb)AAD4488.1 - (AF078856) p47 [Homo sapiens]	Contains protein domain (PF00769) - UBX domain		

2317	87020571 (4633, 4634)			UNCLASSIFIED	22278998, 60432048, 264910, 60432228, 264686, 264687, 264688, 264689, 264558, 18108385
2318	79950879 (4635, 4636)			UNCLASSIFIED	264488, 264569, 18108396, 52646365, 22278994, 22278995, 22278996, 56994075, 35696286, 22278997, 22278998, 264259, 52645080, 28331825, 29331826, 29331827, 29331828, 28331830, 56182435, 60170831, 60432228, 60431735, 33857402, 21908754, 52644288, 87186474, 285011, 87168559, 285017, 265018, 265019, 18108351, 264448, 18108354, 264288, 264368, 52644229, 21908754, 21908785, 21908786, 21908787, 21908788, 21908789, 265021, 265022, 52644150, 33657023, 52645129, 33657109, 27486284, 33657349, 35695783, 18108370, 18108376, 18108379, 35695423, 264558, 83373044, 18108385, 58526486, 87168518, 264554, 264565, 264566
2320	91622426 (4639, 4640)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - III ALU SUBFAMILY SQ WARNING ENTRY !!!!		kinase	22278994, 60432049, 60432289, 29331827, 264511, 265008, 52646317, 265017, 265019, 21908785, 18108372, 18108387, 22279002, 264488, 264687, 18108394, 264689, 21906765, 18108397, 18108398, 21906767, 21906768, 65274791, 22278995, 35695855, 22278998, 285021, 265022, 264510, 265008, 264511, 264512, 265008, 60170615, 264555, 264636, 264556, 18108361, 264259, 60432229, 33657023, 264557, 264558, 264683, 60433356, 264559, 60433438, 29331824, 18108365, 18108384, 29331825, 18108385, 33109554, 29331827, 56526486, 29146499, 265011, 60432113, 285017, 265018, 264508, 264563, 264482, 284508, 18108351, 284448, 284907, 284682, 18108370, 264683, 264908, 264288, 264909, 18108354, 264486, 264567
2321	94320377 (4641, 4642)	Novel Protein sim. GBank gi 3873837 emb CAB02700  - (281028) Similarity to S.pombe hypothetical protein C1D4.09C (SW:Q10154); cDNA EST EMBL: T00543 comes from this gene; cDNA EST EMBL: T01082 comes from this gene; cDNA EST EMBL: T01321 comes from this gene; cDNA EST EMBL: T02288 com...		UNCLASSIFIED	22278996, 264907, 264511, 264757, 18108351, 284788, 284638
2322	87803165 (4643, 4644)	Novel Protein sim. GBank gi 5678957 emb CAB51885.1  - (AL109630) BACR7A4.y [Drosophila melanogaster]	Contains protein domain (PF00105) - short chain dehydrogenase	dehydrogenase	

2323	84840445 (4845, 4846)	Novel Protein sim. GBank gij2494162[sp Q10005 YRY1_CAEEL - HYPOTHETICAL 39.9 KD PROTEIN T15H8.1 IN CHROMOSOME II PRECURSOR	Contains protein domain (PF00226) - leph DnaJ domain	22278994, 22278995, 22278997, 60432049, 264259, 29331822, 33656970, 264509, 56182435, 264511, 265008, 60433356, 60433438, 55812038, 33108954, 21908754, 85688542, 87188474, 265011, 87188559, 265017, 265019, 264760, 264681, 18108351, 264369, 264286, 18108355, 264687, 264688, 21906765, 21806767, 21806768, 55811957, 35695917, 265021, 33657023, 18108362, 27486262, 55811578, 264631, 264555, 83373044, 87188518, 60432113, 22279002
2324	86533607 (4647, 4848)	Novel Protein sim. GBank gij231885[sp P29981 CP4C_BLADI - CYTOCHROME P450 4C1 (CYP1VC1)	ATPase-associated	264592, 264593, 265020
2325	88165074 (4649, 4850)	Novel Protein sim. GBank (AL096732) hypothetical protein [Homo sapiens]	Contains protein domain (PF00067) - Cytb450	265006, 264759, 35695855, 56182323
2326	84390962 (4651, 4652)	Novel Protein sim. GBank gij231885[sp P29981 CP4C_BLADI - CYTOCHROME P450 4C1 (CYP1VC1)	UNCLASSIFIED	
2327	88081848 (4653, 4654)	Novel Protein sim. GBank gij240227[dbj BAA74892.1  - (AB020676) KIAA0869 protein [Homo sapiens]	UNCLASSIFIED	264259, 264508, 264905, 264906, 264907, 264908, 265007, 264512, 264910, 264758, 265010, 264766, 264768, 264769, 33657023, 264693, 264628, 264631, 264634, 264638, 264639, 264488
2328	83388428 (4655, 4656)	Novel Protein sim. GBank gij1245105 (J46463) - glutamine repeat protein-1 [Mus musculus]	UNCLASSIFIED	60433438, 264595, 265017, 264766, 264692, 264629, 264635, 264636, 264638, 56182323, 60432113, 264566
2329	87604478 (4657, 4658)	Novel Protein sim. GBank gij1169343[sp P42209 DIF6_MOUSE - DIFF6 PROTEIN	UNCLASSIFIED	265017, 264685, 60432113, 264088
2330	87335388 (4659, 4660)	Novel Protein sim. GBank gij5679136[gb AAD46874.1 AF16093 - (AF160934)	transport	265009
2331	88980463 (4661, 4662)	Novel Protein sim. GBank gij5679136[gb AAD46874.1 AF16093 - (AF160934)	ATPase-associated	35696286, 22278998, 29331824, 60424269, 265006, 265008, 265018, 264448, 264764, 21906765, 35695917, 35698855, 264636, 22279000, 264586
2332	87784182 (4663, 4664)	Novel Protein sim. GBank gij2104452[emb CAB08779  - (Z95397) unknown [Schizosaccharomyces pombe]	UNCLASSIFIED	56182575, 56994075, 29331826, 29331828, 284107, 33657402, 87168559, 264683, 35695917, 265021, 33657023, 263976
2333	88206958 (4665, 4666)	Novel Protein sim. GBank gij3879985[emb CAA92691.1  - (Z68318) cDNA EST CEMSD62F comes from this gene; cDNA EST EMBL:CO7830 comes from this gene; cDNA EST EMBL:CO9493 comes from this gene; cDNA EST yk415e8.3 comes from this gene; cDNA EST yk415e8.5 comes from this gene; cDNA EST...	UNCLASSIFIED	
2334	94319788 (4667, 4668)	Novel Protein sim. GBank gij4886270[gb AAB52261.2  - (U97002) similar to acyl-CoA dehydrogenases and epoxide hydrolases; Pfam domain PF00441 (Acyl-CoA_dh). Score=57.4, E-value=1.7e-16, N=2; contains similarity to Pfam domain PF00702 (Hydrolase), Score=57.4, E- value=1e-13, N=1 C...	Contains protein domain (PF00441) - dehydrogenase Acyl-CoA dehydrogenase	56182575, 29331825, 21906768, 264636, 83373044

2335	80046103 (4668, 4670)	Novel Protein sim. GBank gll3283350 (AF062378) - calmodulin-binding protein SHA1 [Mus musculus]	Contains protein domain (PF00612) - struct	18108351, 21908769, 264555
2336	95196121 (4671, 4672)	Novel Protein sim. GBank gll1929096 (CAA72805) - (Y12080) putative 3,4-dihydroxy-2-butanone kinase [Lycopersicon esculentum]	kinase	264907, 35695917, 18108379
2337	95345810 (4673, 4674)	Novel Protein sim. GBank gll4495063 (CAB39181.1) - (Z84988) dJ108K11.3 (similar to yeast suppressor protein SRP40) [Homo sapiens]	UNCLASSIFIED	35696286, 22278998, 56182181, 29331825, 60424269, 56182435, 33657402, 55811386, 265017, 265018, 265019, 21906768, 35695917, 264691, 33657023, 33657109, 263872, 35696423, 35695855, 60432113
2338	87634045 (4675, 4676)	Novel Protein sim. GBank gll2224689 (dJ1BAA20829) - (AB002372) KIAA0374 [Homo sapiens]	Contains protein domain (PF00323) - Mammalian defensin	18108394, 29331822, 66714117, 60432289, 264906, 56182435, 265009, 60433438, 264596, 265010, 265019, 18108354, 264288, 264369, 55811957, 265021, 33657023, 263976, 55811576, 264632, 56182323, 264639
2339	85663319 (4677, 4678)	Novel Protein sim. GBank gll3873550 (CAB22127) - (AL033534) serine-rich protein [Schizosaccharomyces pombe]	UNCLASSIFIED	35696286, 264592, 264369, 264691, 264558
2340	90937716 (4679, 4680)			
2341	8775281 (4681, 4682)	Novel Protein sim. GBank gll3874563 (CAB02797) - (Z81042) similar to Yeast hypothetical protein YEF6 like; cDNA EST yk206h5.3 comes from this gene; cDNA EST yk206h5.5 comes from this gene; cDNA EST yk303h1.3 comes from this gene; cDNA EST yk303h1.5 comes from this gene; cDNA ...	kinase	264259, 264908, 264909, 264682, 22279000, 264488, 65274572, 22278995, 22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331825, 66714117, 29331828, 60432289, 35696052, 264905, 264509, 29331830, 265008, 60170831, 60432228, 60433356, 87168474, 265017, 265018, 265019, 284448, 264389, 264288, 21906765, 21906766, 55811957, 35695917, 265020, 265022, 52644150, 33657023, 65274820, 33657109, 18108370, 18108376, 55810764, 35696423, 55811576, 264556, 264558, 18108385, 60432113, 264563, 264564, 264565, 264566, 264567, 264907, 264512, 265011, 284683
2343	8775448 (4685, 4686)	Novel Protein sim. GBank gll492974 (gblAAD34131.1) (AF151894) CGI-136 protein [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat	
2344	79953198 (4687, 4688)	Novel Protein sim. GBank gll2506307 (spP1394) (CA1C, CHICK - COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR (FIBROCHIMERIN)	UNCLASSIFIED	264758
2345	94319799 (4689, 4690)		Contains protein domain (PF00092) - von Willebrand factor type A domain	264488, 264259, 66712502, 264759, 83373044, 264566

2346	84131820 (4691, 4692)	Novel Protein sim. GBank gll255411 (U53153) - one short region of weak similarity to S. cerevisiae protease A inhibitor 3 (SP:P01094) and another short region of weak similarity to S. cerevisiae glucose repression mediator protein (SP:P14822) [Caenorhabditis elegans]	Contains protein domain (PF00515) - TPR Domain	proteaseinhib	35696286, 22278998, 264259, 35696052, 29331828, 33657402, 60433356, 33109854, 87168559, 264603, 265019, 18108351, 264681, 264685, 21908786, 265021, 33657108, 55811578, 35695855, 284637, 52644332, 264557, 83373044, 22279000, 22279002
2347	85330367 (4693, 4694)				22278997, 264511, 264693, 264694, 264768, 264687, 264688, 264691, 264692, 55811576
2348	95196133 (4695, 4696)	Novel Protein sim. GBank gll1929056[emb]CAA728051 - (Y12090) putative 3,4-dihydroxy-2-butanone kinase [Lycopersicon esculentum]		kinase	18108394, 35696286, 264259, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265006, 265007, 264512, 265008, 265009, 264910, 264591, 264592, 264593, 264594, 264757, 264595, 264596, 264758, 265011, 264601, 264782, 18108351, 264764, 264288, 264766, 264768, 264689, 35695917, 264693, 264628, 18108370, 264638, 18108374, 35696423, 264631, 264635, 264636, 264637, 264638, 264639, 83373044, 18108385, 264587, 264486
2349	87776502 (4697, 4698)	Novel Protein sim. GBank gll4884106[emb]CAB43254.1) - (AL050062) hypothetical protein [Homo sapiens]			35696052, 29146499, 264908, 264369
2350	88260594 (4699, 4700)				22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331827, 29331828, 33109954, 21908754, 265010, 87168559, 285018, 265019, 264781, 264681, 264288, 18108357, 21908766, 21908767, 264691, 264682, 35695855, 87168518, 22279000, 22279002, 264482
2351	86888042 (4701, 4702)	Novel Protein sim. GBank gll728832[sp]P39189[ALU2_HUMAN - III] ALU SUBFAMILY SB WARNING ENTRY IIII		kinase	56182575, 264909, 265006, 264558
2352	87337196 (4703, 4704)	Novel Protein sim. GBank gll731637[sp]P38760[YHH5_YEAST - HYPOTHETICAL 75.9 KD PROTEIN IN SPO13-ARG4 INTERGENIC REGION]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264259, 264448
2353	91638784 (4705, 4706)	Novel Protein sim. GBank gll1346855[sp]P48809[RB27_DROME - HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN 27C (HNRNP 48) (HRP48.1)]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	dna_rna_bind	29331826, 55812038, 265019, 264692, 264636
2354	87337199 (4707, 4708)	Novel Protein sim. GBank gll731637[sp]P38760[YHH5_YEAST - HYPOTHETICAL 75.9 KD PROTEIN IN SPO13-ARG4 INTERGENIC REGION]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	29331824, 264908, 265006, 265008

2355	91838788 (4709, 4710)	Novel Protein sim. GBank gll4938503[emb]CAB43861.1] - (AL078465) hnRNP-like protein [Arabidopsis thaliana]	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	- dna_rna_bind	56954075, 22278986, 35696285, 22278999, 284259, 29331825, 29331826, 29331828, 29146498, 284905, 264908, 265006, 264759, 87168474, 285010, 265017, 264687, 21908765, 21908767, 21908769, 264691, 264692, 263987, 18108370, 87168518, 22279000
2356	85327688 (4711, 4712)	Novel Protein sim. GBank gll5138920[gb]AAD40377.1] - (AF092135) PTD014 [Homo sapiens]			52644507, 22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331824, 66714117, 29331825, 60432289, 35696052, 29331828, 264908, 66712502, 264512, 265007, 285008, 60170831, 60432229, 60433358, 60433438, 264758, 52646317, 33109954, 21908764, 55811366, 87168474, 265017, 265018, 264605, 285019, 264681, 264682, 264448, 264368, 264288, 264685, 264768, 21908765, 21908766, 21908767, 21908768, 21908769, 265021, 60170615, 33657109, 27486264, 35695763, 55810764, 18108379, 35696423, 55811576, 35695855, 60170394, 56182323, 83373044, 18108385, 56526486, 264404, 60432113, 22279000, 264482, 264563, 264566, 264486, 264587
2357	87775458 (4713, 4714)	Novel Protein sim. GBank gll4929741[gb]AAD34131.1]AF15189 - (AF151894) CGI-136 protein [Homo sapiens]		UNCLASSIFIED	284488, 264769, 18108394, 264259, 29331822, 18108370, 18108374, 264510, 265017, 264482, 264563, 264762, 264565, 264566, 264369, 18108354
2358	87777078 (4715, 4716)	Novel Protein sim. GBank gll4218005 (AC006135) - putative vicilin storage protein (globulin-like) [Arabidopsis thaliana]		UNCLASSIFIED	22278997, 22278999, 264509, 264905, 264592, 18108351, 264681, 264682, 264769, 32833986, 18108374, 264566, 18108385, 264482
2359	87755859 (4717, 4718)	Novel Protein sim. GBank gll1086830 (U41264) - coded for by C. elegans cDNA yk2018.5; coded for by C. elegans cDNA yk4491.5; coded for by C. elegans cDNA yk1267.5; coded for by C. elegans cDNA yk3596.5; coded for by C. elegans cDNA yk2018.5; coded for by C. elegans cDNA yk16912 ...		UNCLASSIFIED	35696288, 22278998, 264905, 264511, 265007, 265008, 60433438, 284288, 284686, 21908769, 265020, 284692, 35696555, 264558, 56526486, 264563
2360	80046125 (4719, 4720)	Novel Protein sim. GBank gll3881545[emb]CAA93779 - (Z69904) cDNA EST yk428d5.3 comes from this gene; cDNA EST yk428d5.5 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	22278997, 29331826, 263981, 22279000
2361	94232191 (4721, 4722)	Novel Protein sim. GBank gll746487 (U23514) - No definition line found [Caenorhabditis elegans]			22278995, 22278999, 264512, 265009, 264757, 21908765, 65274620, 18108370, 60431528, 18108374, 264635, 60170394, 264482

2362	81721193 (4723, 4724)	Novel Protein sim. GBank gii1171083jip19706jmy58.ACACA - MYOSIN HEAVY CHAIN IB (MYOSIN HEAVY CHAIN IL)		UNCLASSIFIED	22278998, 264259, 29331822, 29331824, 60432289, 264509, 264512, 60432229, 60433356, 264448, 264682, 264683, 264369, 21906765, 21906768, 21906769, 60432113, 22279000, 22279002
2363	95006635 (4725, 4726)	Novel Protein sim. GBank gii854065jemb(CAA58337) - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264907, 264629, 264635
2364	84827104 (4727, 4728)	Novel Protein sim. GBank gii5639830jipAAD45888.1(AF 14601) - (AF 146018) hydroxypyruvate reductase [Homo sapiens]	Contains protein domain (PF00389) - D-isomer specific 2-hydroxyacid dehydrogenases	reductase	264488, 18108394, 264887, 18108398, 22278998, 56994075, 35696286, 22278997, 22278998, 264239, 66714117, 29331825, 35696052, 264509, 264505, 264506, 264807, 264908, 66712502, 264509, 264511, 265006, 264512, 265007, 265008, 33657402, 264758, 21906754, 87168474, 265010, 87168559, 264603, 265017, 265018, 265019, 264780, 264782, 18108351, 264448, 264784, 264683, 264684, 264288, 18108355, 264786, 18108358, 264689, 18108359, 21906765, 21906766, 21906767, 35695917, 265020, 265021, 265022, 60170615, 52644150, 264691, 33657023, 264692, 18108364, 33657109, 18108368, 18108370, 18108374, 35698423, 35695855, 264635, 264558, 264557, 264639, 60170394, 83373044, 18108383, 18108384, 18108385, 18108388, 56526489, 264482, 264564, 264486
2365	94140746 (4729, 4730)	Novel Protein sim. GBank gii1840045 (U49062) - transporter protein [Homo sapiens]		transport	22278998, 22278998, 22278999, 264907, 264909, 264910, 33657402, 264758, 264600, 264766, 264887, 264889, 21906765, 21906767, 21906768, 21906769, 265021, 33657023, 33657109, 83373044, 264566
2366	94312388 (4731, 4732)			UNCLASSIFIED	52644507, 264259, 29331824, 29331825, 29331826, 29331827, 29331828, 264907, 29331830, 264809, 264511, 265008, 33657402, 264595, 52646317, 265017, 265018, 265019, 264605, 264685, 264766, 264689, 21906766, 21906769, 35695917, 265020, 265021, 265022, 52644150, 35695855, 52644332, 18108385, 18108387, 264564, 264568
2367	94140910 (4733, 4734)	Novel Protein sim. GBank gii1065457 (U40410) - C54G7.4 gene product [Caenorhabditis elegans]	Contains protein domain (PF00400) - IWD domain, G-beta repeat	UNCLASSIFIED	35696286, 21906768, 55810764, 65274791, 264587
2368	94322190 (4735, 4736)				264828

2369	94314334 (4737, 4738)	Novel Protein sim. GBank gll5360901[dbj]BA02158.1] - (AB029343) a-helix coiled-coil rod homologue [Homo sapiens]		struct	52644507, 52646842, 35698286, 264092, 264094, 52645080, 35698052, 264107, 29331830, 52644045, 265006, 265007, 265009, 52644296, 52644229, 264689, 21908765, 21908766, 35695917, 265020, 52644150, 263967, 33657109, 27486265, 35695763, 18108370, 263974, 18108374, 18108376, 52644332, 263981, 18108385, 264508, 264909, 264596
2370	79804120 (4739, 4740)			UNCLASSIFIED	264389
2371	57280406 (4741, 4742)			UNCLASSIFIED	263967, 263981
2372	87642413 (4743, 4744)			UNCLASSIFIED	29331826, 265010, 265019, 35695917, 264634, 60432113
2373	87418611 (4745, 4746)	Novel Protein sim. GBank gll4589582[dbj]BA076813.1] - (AB023186) KIAA0969 protein [Homo sapiens]		UNCLASSIFIED	265008, 265007, 265008, 265009, 265011, 264766, 35695917, 35695855, 263981, 264557, 264585
2374	94123665 (4747, 4748)	Novel Protein sim. GBank gll5105131[dbj]BA00445.1] - (AP000061) 24Gaa long hypothetical ribonuclease PH [Aeropyrum pernix]	Contains protein domain (PF01138) - 3' exoribonuclease family	UNCLASSIFIED	60432049, 29331824, 264807, 52644045, 264512, 60433356, 21908754, 52644298, 87168559, 264448, 21908765, 21908768, 21908769, 33657023, 18108368, 55811576, 52644332
2375	87731355 (4749, 4750)	Novel Protein sim. GBank gll1351115[dbj]P47758[SRPB_MOUSE - SIGNAL RECOGNITION PARTICLE RECEPTOR BETA SUBUNIT (SR-BETA)]		UNCLASSIFIED	264259, 29331830, 264909, 264910, 265009, 60433438, 21908754, 285017, 285018, 265019, 264682, 264288, 264685, 21908767, 263972, 35695855, 87168518, 60432113
2376	87613744 (4751, 4752)	Novel Protein sim. GBank gll2645435 (AF007780) - CHD3 [Drosophila melanogaster]	Contains protein domain (PF00628) - PHD-finger	UNCLASSIFIED	18108394, 85274572, 22278997, 22278999, 284095, 29331822, 29147620, 29331824, 66714117, 29331825, 29331826, 29331828, 33656970, 28146498, 29146499, 284509, 265006, 265007, 265008, 265009, 60170831, 265010, 265011, 265018, 55811150, 18108351, 264764, 264288, 21908767, 21908768, 29148627, 29148629, 265021, 33657023, 33657109, 18108370, 18108374, 18108379, 35696423, 264556, 83373044, 18108385, 18108386, 56526486, 22279000, 22279002, 264583
2377	95319689 (4753, 4754)	Novel Protein sim. GBank gll5257005[gb]AAD41239.1] - (AF083249) Rb binding protein homolog [Homo sapiens]	Contains protein domain (PF01368) - ARID DNA binding domain	UNCLASSIFIED	65274572, 56182375, 35696286, 264259, 29331822, 29331824, 66714117, 29331825, 29331826, 60432289, 29331827, 56182435, 264510, 265009, 60433358, 87168474, 265011, 265018, 264288, 21906765, 33657023, 264557, 56182323, 83373044, 18108385, 22279002, 264482, 265017, 264288, 21906768
2378	94137032 (4755, 4756)	Novel Protein sim. GBank gll1072198 (U40942) - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	
2379	65444324 (4757, 4758)	Novel Protein sim. GBank gll3337357 (AC004481) - Hypothetical protein [Arabidopsis thaliana]	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	



2380	88923062 (4758, 4760)	Novel Protein sim. GBank gi 4502939 ref NP_001845.1 pCOL1 - collagen, type XI, alpha 1	Contains protein domain (PF01410) - collagen Fibrillar collagen C-terminal domain	264908, 264910, 265011
2381	87808241 (4761, 4762)	Novel Protein sim. GBank gi 445560 jmbi CAB36555  - (AL031846) dJ742C18.5 (novel Chromobox protein) [Homo sapiens]	Contains protein domain (PF00385) - helicase 'chromo' (C-HRromatin Organization Modifier) domain	56182575, 264091, 264093, 264259, 29331825, 264105, 264906, 60433356, 21905754, 265017, 265019, 264683, 264288, 264685, 264686, 264687, 264691, 264692, 264693, 55811576, 264636, 264567
2382	81225982 (4763, 4764)	Novel Protein sim. GBank gi 4325130 gb AAD17276  - (AF119716) dMi-2 protein [Drosophila melanogaster]	Contains protein domain (PF00628) - transport PHD-finger	29331824, 60432288, 264605, 264596, 21905754, 264769, 265022, 264693, 263987, 33657109, 264629, 264631, 264558, 83373044, 60432113, 264482
2383	87442841 (4765, 4766)	Novel Protein sim. GBank gi 1902882 db BAA19005  - (D89049) lectin-like oxidized LDL receptor [Bos taurus]	Contains protein domain (PF00059) - glycoprotein Lectin C-type domain	265009, 21908765, 21908768
2384	95354766 (4767, 4768)	Novel Protein sim. GBank gi 2462851  (AF016252) - Spinophilin [Rattus norvegicus]	Contains protein domain (PF00595) - struct PDZ domain (Also known as DHR or GLGF).	35696286, 22278998, 52645080, 29331824, 29331826, 35698052, 29331828, 264908, 264828, 52644045, 265006, 265008, 265009, 33109954, 33657084, 52644296, 265011, 265017, 265018, 264683, 52644229, 21908765, 21908767, 21906768, 265020, 52644150, 33657023, 264693, 65274620, 52645128, 33657109, 33657182, 27486261, 27486282, 27486264, 33657349, 27486285, 35695763, 18108374, 35695855, 264634, 264555, 264556, 264557, 52644332, 264558, 264559, 18108385, 22279000, 22279002, 264486, 52644507, 52645156, 264887
2385	95419485 (4769, 4770)		UNCLASSIFIED	52646365, 22278995, 22278996, 22278997, 22278999, 264259, 52645080, 29331822, 29331826, 35696052, 52644045, 265006, 265007, 265008, 265009, 264910, 60432229, 60433356, 52646317, 21906754, 265019, 264448, 264683, 264686, 264687, 264689, 21906765, 21906766, 21906767, 21906769, 55811937, 265021, 265022, 264690, 264691, 264692, 65274620, 33657109, 18108370, 264631, 52644332, 22279000, 22279002, 264583, 264585, 264587
2386	94742648 (4771, 4772)	Novel Protein sim. GBank gi 4928888 gb AAD34110.1 AF15187 - (AF151873) CGI-115 protein [Homo sapiens]	glycoprotein	264488, 22278995, 22278996, 22278997, 264259, 29146488, 264112, 264511, 60170831, 60432229, 264595, 60433438, 87168474, 87168559, 264682, 21906765, 21906768, 21906767, 21906769, 29148629, 35695917, 265021, 264690, 33657109, 264628, 18108376, 83373044, 60432113, 22279000, 264564, 264566, 264487

2387	14997990 (4773, 4774)			UNCLASSIFIED	254634
2388	11424604 (4775, 4776)			UNCLASSIFIED	254595
2389	95310650 (4777, 4778)	Novel Protein sim. GBank gi4758058 NP_004372.1 pCREB - cAMP responsive element binding protein-like 1	Contains protein domain (PF00170) - bZIP transcription factor	dna_rna_bind	284488, 22278998, 22278999, 264509, 264905, 264906, 264907, 264908, 264909, 285008, 284511, 284512, 284910, 284591, 21908754, 264601, 264604, 264761, 18108351, 264764, 264288, 264766, 264768, 264769, 21906765, 21906768, 264892, 264693, 35696423, 264635, 264638, 264555, 83373044, 22278000, 264488
2390	94320912 (4779, 4780)	Novel Protein sim. GBank gi1644239 d BAA12223  - (D84103) mitochondrial DNA polymerase gamma [Homo sapiens]	Contains protein domain (PF00476) - DNA polymerase family A	polymerase	52644507, 56182575, 22278995, 35696286, 22278996, 22278997, 22278999, 28331822, 28331825, 28331826, 35696052, 264905, 52644045, 265009, 264758, 264759, 33109954, 52644296, 85658542, 285011, 265017, 265018, 264605, 52644229, 21906765, 21906767, 21906768, 21908769, 35695917, 52644150, 33657023, 33657109, 33657349, 35695763, 18108370, 18108374, 18108376, 35696423, 35695855, 264555, 52644332, 56182323, 60170394, 83373044, 56526486
2391	80035194 (4781, 4782)			UNCLASSIFIED	263376
2392	84245016 (4783, 4784)	Novel Protein sim. GBank gi4240169 d BAA74853.1  - (AB020647) KIAA0840 protein [Homo sapiens]	Contains protein domain (PF00560) - Leucine Rich Repeat	nuclease	35696286, 35696052, 284508, 264905, 264509, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 264910, 265009, 264591, 264758, 264600, 264604, 264762, 264448, 264764, 264369, 264766, 264768, 264759, 264689, 35695917, 264629, 18108374, 283978, 35696423, 35695855, 264631, 264634, 264635, 264638, 264637, 264638, 60170394, 264638, 264565, 264486

2393	85302633 (4785, 4786)	Novel Protein sim. GBank gi 4506667 ref NP_000993.1 PRPLP - ribosomal protein, large, P0	Contains protein domain (PF00466) - ribosomal prot Ribosomal protein L10	18108392, 60424178, 284489, 18108394, 18108397, 22278995, 56994075, 35696286, 22278996, 22278997, 22278998, 264093, 60432049, 264259, 29331822, 29147620, 20281099, 29331824, 29331825, 68714117, 60432289, 29331826, 29331827, 29331828, 35696052, 28148498, 284508, 284509, 264905, 264907, 264908, 68712502, 52644045, 264828, 264909, 56182435, 264112, 284113, 284510, 265006, 264511, 265007, 265008, 265009, 264910, 264591, 264593, 60433356, 264595, 60433438, 52646317, 33109954, 21906754, 55811386, 265010, 265011, 265017, 265018, 265019, 264681, 264762, 18108351, 264763, 284682, 264784, 264683, 284369, 284288, 18108354, 264766, 264686, 284687, 264688, 284689, 18108359, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 29146629, 29148784, 35695917, 265021, 265022, 33657023, 264692, 264693, 18108364, 33657109, 18108366, 27486261, 27486282, 33657348, 35695763, 18108370, 283972, 264629, 18108374, 263977, 18108378, 263978, 55810764, 35696423, 35696555, 264634, 60431850, 284555, 284637, 264557, 263981, 264558, 18108381, 60170394, 35696286, 22278997, 22278998, 56182181, 35696052, 265006, 264592, 55811386, 265010, 265011, 265017, 265019, 264448, 284683, 284288, 21906765, 21906768, 21908789, 55811957, 35695917, 33657023, 65274820, 33657182, 33657349, 35695763, 18108374, 18108376, 55810764, 55811578, 35696423, 60170394, 18108385, 264584, 264556, 264567
2384	84323266 (4787, 4788)	Novel Protein sim. GBank gi 4159888 (AC004908) - zinc finger protein from gene of uncertain exon structure; similar to Q99676 (PID.g3025333) [Homo sapiens]	Contains protein domain (PF00098) - Zinc finger, C2H2 type	dna_ma_bind
2395	85287212 (4789, 4790)	Novel Protein sim. GBank gi 5712756 gb AA047636.1 AF16079 - (AF16079) calcium transporter CaT1 [Rattus norvegicus]		dna_ma_bind

2398	95096700 (4791, 4792)	Novel Protein sim. GBank gij106322[pilB34087 - hypothetical protein (L1H 3' region) - human	Contains protein domain (PF00560) - Leucine Rich Repeat	nuclease	52646385, 18108387, 56182575, 35696286, 22278997, 22278999, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264108, 264907, 29331830, 68712502, 264110, 60170831, 264591, 33657402, 60433438, 55812038, 33109954, 21908754, 33857094, 87168474, 265017, 265018, 265019, 264760, 264448, 264288, 264766, 52644229, 21908766, 21908767, 285020, 265021, 60170815, 264692, 33857023, 65274620, 52645128, 33857182, 27486262, 27486264, 27486265, 264629, 18108374, 35896423, 35895855, 264631, 264556, 52644332, 264558, 83373044, 18108388, 87166518, 22279002, 264482, 52644507, 52045156, 56182575, 264259, 29147620, 264905, 264907, 264908, 264809, 264910, 264758, 52644296, 264803, 264804, 264762, 264681, 264764, 18108357, 264769, 21908768, 264693, 264628, 264635, 264638, 264639, 264564
2397	87280854 (4793, 4794)				UNCLASSIFIED
2398	88047689 (4795, 4796)	Novel Protein sim. GBank gij3258609 (AC005178) - H53 GST1 (Homo sapiens)			UNCLASSIFIED
2399	87738365 (4797, 4798)	Novel Protein sim. GBank gij786117 (L41834) - nuclear protein (Ensis minor)			56994075, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 56714117, 60432289, 264906, 29331830, 56182435, 264112, 264910, 33109954, 21908754, 87168474, 264600, 265017, 265018, 265019, 264764, 264785, 21908785, 21908786, 265022, 60170815, 33857023, 18108370, 21908767, 21908769, 35895917, 265020, 265022, 60170815, 33857023, 18108370, 18108374, 264556, 60170394, 264558, 87168518, 22279000, 22279002, 264564, 264566, 264487
2400	91214116 (4799, 4800)	Novel Protein sim. GBank gij2352822[gljAAB69285.1] - (AF008945) glucose-6-phosphate [Haplochromis nubilus]		phosphatase	21908786, 52844296, 265017, 27486282, 52844296, 265017

2401	91214118 (4801, 4802)	Novel Protein sim. GBank gij2352822[gijAA669285.1] - (AF008945) glucose-6-phosphatase [Haplochromis nubilus]		phosphatase	52644507, 52645156, 52644228, 264688, 21908764, 21908765, 52646385, 52646842, 21906766, 21906767, 21906768, 22278995, 35695917, 56994075, 35696286, 22278998, 22278997, 265020, 22278998, 22278999, 264259, 33657023, 52645080, 264693, 29331824, 33657109, 52645129, 26331826, 33657182, 29331827, 35696052, 27486261, 27486262, 33656970, 33657349, 27486265, 35695763, 264106, 264905, 35696423, 35695855, 265006, 265007, 265008, 265009, 264637, 52644332, 55812038, 52646317, 18108385, 52644288, 87168474, 265010, 87168559, 80432113, 265017, 265018, 265019, 264583, 264288
2402	91221408 (4803, 4804)	Novel Protein sim. GBank gij4689258[gijAAD27832.1]AF12185 - (AF121859) sorting nexin 8 [Homo sapiens]			264907, 264908, 264909, 264566
2403	84135432 (4805, 4806)	Novel Protein sim. GBank gij4929575[gijAAD34048.1]AF15181 - (AF151811) CGI-53 protein [Homo sapiens]	Contains protein domain (PF00062) - C-type lysozyme/alpha-lactalbumin family		22278999, 35696052, 265018, 264688, 264693, 83373044, 264567
2404	95312605 (4807, 4808)	Novel Protein sim. GBank gij2315798 (AF016685) - similar to short chain-type dehydrogenases [Caenorhabditis elegans]	Contains protein domain (PF00106) - short chain dehydrogenase	dehydrogenase	35696286, 29331826, 35696052, 265008, 265018, 21906769, 264564
2405	84311851 (4808, 4810)	Novel Protein sim. GBank gij464178[gijBAA03581] - (D14853) polyprotein [Hepatitis C virus]		UNCLASSIFIED	35696286, 29331822, 265007, 21906754, 265017, 265018, 265019, 264763, 264369, 21906765, 35695917, 265020, 265021, 52644150, 264693, 35695855, 264632, 52644332, 22279002
2406	88094501 (4811, 4812)	Novel Protein sim. GBank gij2773363 (AF041382) - microtubule binding protein D-CLIP-190 [Drosophila melanogaster]	Contains protein domain (PF01302) - CAP-Gly domain	struct	52646842, 22278994, 22278996, 35696286, 22278997, 22278998, 22278999, 264092, 264093, 60432049, 264259, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 264102, 264106, 264906, 52644045, 265007, 265008, 265009, 264910, 284592, 80433398, 60433438, 33109954, 265010, 265011, 265018, 265019, 264369, 264685, 264686, 21906768, 21906769, 52644150, 264693, 52645129, 264628, 35696423, 264632, 56182323, 264639, 22278000, 22278002, 264583
2407	79465005 (4813, 4814)			UNCLASSIFIED	264685, 264686
2408	87391503 (4815, 4816)	Novel Protein sim. GBank gij423442[gijS33513 - gene Fif protein - mouse]		UNCLASSIFIED	264910, 265010, 264448, 264557

2409	94741770 (4817, 4818)	Novel Protein sim. GBank gjl1176601sp45966YNZ6_CAEEL - HYPOTHETICAL 20.8 KD PROTEIN T09A5.6 IN CHROMOSOME III	UNCLASSIFIED	22278995, 22278996, 22278997, 264097, 29331822, 29331824, 29331827, 29146498, 52644045, 60433438, 33857084, 87168474, 264780, 21906787, 29148627, 29148629, 52644150, 33657023, 263967, 20281069, 18108374, 20281071, 56182323, 83373044, 18108385, 87168518 22278998, 264259
2410	87604860 (4819, 4820)	Novel Protein sim. GBank gjl5986262jbiAAC48032.21 - (U64849) Contains similarity to Plam domain; PF00646 (F- box). Score=28.7, E-value=4.3e-05, N=1 [Caenorhabditis elegans]	Contains protein domain (PF00080) - Copper/zinc superoxide dismutase (SODC)	
2411	87534633 (4821, 4822)	Novel Protein sim. GBank gjl3114713 (AF061346) - Ecdp1 protein [Mus musculus]	Inf	29331824, 29331827, 29331828, 264784, 264389, 33657109, 56182323
2412	87778332 (4823, 4824)	Novel Protein sim. GBank gjl5410336jbiAAD43038.11 - (AF106685) myelin gene expression factor 2 [Homo sapiens]	dna_rna_bind Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	265017, 265018, 265019, 18108351, 21906766, 265020, 33657109, 264559, 18108385
2413	94133820 (4825, 4826)	Novel Protein sim. GBank gjl5262705jembICAB45778.11 - (AL080214) hypothetical protein [Homo sapiens]	struct Contains protein domain (PF00038) - Intermediate filament proteins	264488, 264259, 29331826, 264508, 264805, 264509, 264806, 264807, 264808, 264510, 264511, 264512, 265008, 265009, 264910, 265011, 264682, 264764, 264766, 264686, 264788, 264689, 265021, 33657023, 18108370, 264628, 35693855, 264632, 264634, 264635, 264636, 83373044, 264563, 264584, 264585, 264586, 264587, 264486
2414	94312590 (4827, 4828)	Novel Protein sim. GBank gjl1082340jpirIS52863 - DNA- binding protein R kappa B - human	ubiquitin	52645156, 52646842, 52646385, 18108398, 56182575, 22278994, 22278995, 56994075, 22278996, 35696286, 22278997, 22278998, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 33656970, 29331830, 264808, 60433356, 33657402, 264594, 52646317, 21906754, 33657084, 265010, 87168559, 265017, 265018, 265019, 264369, 264684, 264687, 264688, 56181562, 21906764, 264689, 21906765, 21906766, 21906767, 29148627, 21906769, 265020, 265021, 60170615, 33657023, 264693, 52645129, 33657109, 33657162, 27486261, 27486262, 27486265, 33657349, 18108370, 60431528, 264629, 18108374, 18108376, 55810784, 264636, 52644332, 264638, 264558, 56182323, 83373044, 18108385, 87168518, 22279002

2415	88080002 (4829, 4830)	Novel Protein sim. GBank gij423915 pir A45439 - myosin I heavy chain - rat	Contains protein domain (PF00063) - struct	Myosin head (motor domain)	264259, 264908, 60433356, 33657402, 21906754, 265018, 264687, 264689, 21906769, 55811957, 265021, 264690, 264691, 33657023, 264693, 35696423, 56182323, 56526486
2416	94118336 (4831, 4832)	Novel Protein sim. GBank gij3025445 (AC004528) - R32184_1 [Homo sapiens]		Contains protein domain (PF00153) -	264638
2417	87733334 (4833, 4834)	Novel Protein sim. GBank gij1084944 pir S54495 - hypothetical protein YPR021c - yeast (Saccharomyces cerevisiae)		Mitochondrial carrier proteins	264094, 29331822, 29331824, 29331827, 264369
2418	94234349 (4835, 4836)	Novel Protein sim. GBank gij1176572 sp P45895 YNA4_CAEEL - HYPOTHETICAL 91.0 KD PROTEIN PAR2.4 IN CHROMOSOME III	UNCLASSIFIED	Contains protein domain (PF00411) - Ribosomal protein S11	56994075, 264091, 264259, 29331824, 29331825, 60432288, 29331828, 264905, 264907, 264511, 255009, 60432229, 21906754, 87168558, 265019, 264682, 21906768, 21906769, 265020, 265021, 33657023, 65274620, 18108370, 55811576, 264634, 60170394, 18108385, 22279000, 22279002, 264568
2419	82374249 (4837, 4838)	Novel Protein sim. GBank gij284006 pir S18732 - autoantigen, 84K - human	struct		264569, 264782, 264448, 264691, 264631, 264634, 264555, 264558, 264638, 264558
2420	94844244 (4839, 4840)	Novel Protein sim. GBank gij107621 pir S50755 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii	UNCLASSIFIED		29331824, 29331825, 29331828, 60432229, 33109994, 85658542, 87168474, 265018, 264288, 265020, 264564
2421	87805345 (4841, 4842)		UNCLASSIFIED		264909, 264768, 264638
2422	88084714 (4843, 4844)	Novel Protein sim. GBank gij222456 tbl BAA20772  (AB002311) KIAA0313 [Homo sapiens]	transport	Contains protein domain (PF00617) - RasGEF domain	18108392, 18108394, 18108398, 264906, 265008, 265010, 18108351, 18108374, 18108385
2423	88058390 (4845, 4846)	Novel Protein sim. GBank gij4505153 ref NP_002392.1 pMEKK - MAP/ERK kinase kinase 3	kinase	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	264259, 60432049, 29331822, 29331826, 60432289, 29331828, 265008, 265009, 60433356, 21906754, 265017, 265018, 265019, 21906766, 21906768, 21906769, 265020, 265021, 20281149, 263971, 60432113
2424	94854047 (4847, 4848)	Novel Protein sim. GBank gij2988398 (AC004391) - Unknown gene product [Homo sapiens]	UNCLASSIFIED		56182575, 35696286, 22278997, 60432049, 264259, 29331826, 29331828, 264905, 68712502, 29331830, 60433356, 265011, 265019, 264766, 21906768, 55811957, 264692, 33657023, 33657109, 55811576, 56182323, 83373044, 18108385, 18108388, 60432113, 22279000
2425	87415881 (4849, 4850)	Novel Protein sim. GBank gij2077832 dbj BAA19879  (D85558) Protein Kinase [Rattus norvegicus]	kinase	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	264634
2426	87613945 (4851, 4852)	Novel Protein sim. GBank gij2039368 gb AAB53003.1  (U94619) circulating cathodic antigen [Schistosoma mansoni]	UNCLASSIFIED		22278998, 22278998, 264259, 264102, 264512, 265008, 21908767, 18108370, 18108374, 263976

2427	87822693 (4853, 4854)	Novel Protein sim. GBank gll4680695jdb/AAD27737.1(AF13296 - (AF132962) CGI-28 protein [Homo sapiens])	Contains protein domain (PF00573) - Ribosomal protein L4/L1 family	ribosomalprot	264259, 20281099, 35666052, 265008, 264594, 265011, 264760, 18108351, 264682, 264683, 264369, 264684, 264686, 264687, 264689, 21908766, 264691, 264692, 18108374, 18108377, 264557, 264639, 18108385
2428	85732889 (4855, 4856)	Novel Protein sim. GBank gll1537070 (U63840) - nucleoporin p54 [Rattus norvegicus]			22278998, 22278999, 35666052, 21908754, 264288, 21908765, 21908768, 21908769, 35695917, 265020, 263972, 22279002
2429	87769276 (4857, 4858)	Novel Protein sim. GBank gll601931 (M84316) - neurofilament-H [Oryctolagus cuniculus]	Contains protein domain (PF00711) - Beta defensins	UNCLASSIFIED	22278999, 29331824, 264606, 264609, 264511, 265008, 21908754, 265017, 265018, 285019, 264448, 264683, 264288, 21908765, 21908768, 265021, 264693, 18108381
2430	85948827 (4859, 4860)	Novel Protein sim. GBank gll3860729jemb/CAA14630) - (AJ235270) CELL DIVISION PROTEIN FTSJ (ftsJ)	Contains protein domain (PF01728) - FtsJ cell division protein	UNCLASSIFIED	264112, 264691
2431	87848884 (4861, 4862)	Novel Protein sim. GBank gll3876367jemb/CAA93287) - (Z69360) Weak similarity to Eimeria thrombospondin (PIR Acc. No. A45517); cDNA EST EMBL:M89268 comes from this gene; cDNA EST yk29559.5 comes from this gene [Caenorhabditis elegans]		protease	29331826, 29331827, 35666052, 29146499, 264905, 264908, 264681, 264288, 264689, 21908765, 264692, 35698423
2432	80083033 (4863, 4864)	Novel Protein sim. GBank gll2224593jdbjBAA20784) - (AB002324) KIAA0326 [Homo sapiens]	Contains protein domain (PF00008) - Zinc finger, C2H2 type	transcriptfactor	264569, 284805, 265018, 264762, 264683, 264691, 264558, 264557, 264639, 264558 264563
2433	80055092 (4865, 4866)	Novel Protein sim. GBank gll1263289 (U47859) - fibron-4 [Araneus diadematus]		UNCLASSIFIED	264555
2434	19520148 (4867, 4868)	Novel Protein sim. GBank gll3641352 (AF091234) - putative transcription factor [Mus musculus]		UNCLASSIFIED	265008, 264758, 265010, 264689, 27486261, 263972, 18108374, 18108381
2435	20759044 (4869, 4870)	Novel Protein sim. GBank gll3660014 (AF091088) - unknown [Homo sapiens]	Contains protein domain (PF01256) - Uncharacterized protein family UPF0031	UNCLASSIFIED	29331828, 265007, 265009, 265017, 264760, 264685, 264693, 264565
2436	88044008 (4871, 4872)				29331826, 264508, 264805, 264509, 264806, 264907, 264908, 264909, 264511, 265008, 264512, 264910, 265008, 264591, 33857402, 21908754, 265011, 264760, 264764, 264685, 264686, 264768, 35695917, 33857023, 264693, 264631, 264632, 56182323, 264558, 83373044, 264563, 264564, 264565, 264566, 264567



2439	94850650 (4877, 4878)	Novel Protein sim. GBank gll4263519[gb]AAD15345] - (AC004044) small nuclear riboprotein Sm-D1 (Arabidopsis thaliana)	Contains protein domain (PF01423) - UNCLASSIFIED Sm protein	60424178, 18108397, 56182575, 56181686, 56984075, 22278996, 35698286, 22278997, 22278999, 264259, 52645080, 29331822, 56182181, 29331824, 60424269, 66714117, 29331825, 60432289, 29331826, 29331827, 29331828, 35698052, 29146498, 264509, 264905, 264906, 52644045, 60431735, 3310954, 21806764, 33557084, 55811386, 52644296, 87168474, 265017, 265018, 265019, 18108351, 264448, 264288, 264768, 52644229, 56181582, 21906764, 21906765, 21906768, 21906767, 21906768, 21906769, 35695917, 33657023, 33657108, 33657182, 27486282, 27486284, 33657348, 27486285, 35695763, 18108370, 60431528, 263977, 55810784, 35696423, 65274791, 35895855, 60431850, 56182323, 60432113, 22279000, 22279002, 264567
2440	87641733 (4879, 4880)		UNCLASSIFIED	29331822, 29331824, 29331825, 29331827, 34696032, 264508, 264907, 264510, 265018, 265019, 264448, 264368, 265020, 265021, 56182323, 264639, 22279002
2441	87623914 (4881, 4882)	Novel Protein sim. GBank gll3024889[sp]P56524[Y288_HUMAN - HYPOTHETICAL PROTEIN KIAA0288 (HA6116)	UNCLASSIFIED	264488, 264628, 18108374, 264564
2442	87273590 (4883, 4884)	Novel Protein sim. GBank gll4506013[ref]NP_002703.1[pppp1 - protein phosphatase 1, regulatory subunit 7	Contains protein domain (PF00560) - UNCLASSIFIED Leucine Rich Repeat	22278996, 22278999, 29331822, 264768, 264693
2443	84305949 (4885, 4886)	Novel Protein sim. GBank gll170658[sp]Q02075[KID1_RAT] - RENAL TRANSCRIPTION FACTOR KID-1 (TRANSCRIPTION FACTOR 17)	Contains protein domain (PF01352) - UNCLASSIFIED KRAB box	264908
2444	86086345 (4887, 4888)	Novel Protein sim. GBank gll4758924[ref]NP_004280.1[pNRF3 - nuclear factor (erythroid-derived 2) like 3	Contains protein domain (PF00170) - UNCLASSIFIED bZIP transcription factor	264259, 18108382, 18108383, 18108385, 22279000
2445	87338636 (4889, 4890)	Novel Protein sim. GBank gll2135950[pr][S58222 - PQ-rich protein - human		264259, 35696052, 264369, 18108361
2446	88058293 (4891, 4892)	Novel Protein sim. GBank gll4753887[emb]CAA05409.2] - (A_002424) p65 protein [Rattus norvegicus]	Contains protein domain (PF00095) - UNCLASSIFIED WAP-type (Whey Acidic Protein) 'four-disulfide core'	265011, 264689, 33657023, 263981, 18108385
2447	94845149 (4893, 4894)	Novel Protein sim. GBank gll4885613[ref]NP_005409.1[pST5] - suppression of tumorigenicity 5	cadherin	56182575, 264259, 29331824, 29331825, 29331827, 60433356, 50433438, 264758, 265018, 264682, 65274620, 60431528, 65274791, 56182323

2448	87749680 (4895, 4896)				UNCLASSIFIED	22278996, 22278997, 22278998, 28331826, 3586052, 264107, 264110, 87168474, 87168559, 18108351, 21906767, 21906769, 27486262, 263976
2449	87869075 (4897, 4898)	Novel Protein sim. GBank gil728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII			cadherin	264259, 284828, 265007, 264595, 265021, 56526486
2450	86597784 (4899, 4900)	Novel Protein sim. GBank			UNCLASSIFIED	264906
2451	91014563 (4901, 4902)	gil1710021 sp P35290 RB24_MOUSE - RAS-RELATED PROTEIN RAB-24 (RAB-16)	Contains protein domain (PF00071) - Ras family		glycoprotein	264083, 28331822, 29331824, 29331825, 66714117, 28331826, 28331828, 3586052, 264907, 66712502, 28331830, 284810, 265009, 284758, 265017, 265018, 284782, 264446, 264288, 21908787, 285021, 33657023, 264693, 33657109, 263969, 83373044, 18108385
2452	91230509 (4903, 4904)	Novel Protein sim. GBank gil1504034 db BAA13216  - [D88980] KIAA0227 [Homo sapiens]			isomerase	264102, 264112, 284688, 263972, 18108374, 83373044, 264563
2453	84201088 (4905, 4906)	Novel Protein sim. GBank gil2880079 (AC004142) - similar to murine leucine-rich repeat protein; possible role in neural development by protein-protein interactions; 93% similarity to D49802 (PID:g1369906) [Homo sapiens]	Contains protein domain (PF00560) - Leucine Rich Repeat		ngfirecep	264509, 264512, 18108385
2454	95310691 (4907, 4908)	Novel Protein sim. GBank gil1076802 pi S49915 - extensin like protein - maize	Contains protein domain (PF00170) - bZIP transcription factor		UNCLASSIFIED	263994, 66714117, 29331827, 264508, 264509, 284905, 264906, 264907, 264908, 264908, 264510, 264511, 264512, 265009, 264910, 284591, 284758, 264759, 285010, 265011, 284603, 284604, 284780, 284781, 264762, 18108351, 264764, 284765, 264766, 264686, 264768, 264769, 264534, 264691, 264692, 33657023, 264693, 33657109, 284628, 263978, 35695855, 284634, 264635, 264637, 284638, 264639, 83373044, 18108385, 284583, 284594, 284486
2455	95288301 (4909, 4910)	Novel Protein sim. GBank gil543817 sp P35585 AP47_MOUSE - CLATHRIN COAT ASSEMBLY PROTEIN AP47 (CLATHRIN COAT ASSOCIATED PROTEIN AP47) (GOLGI ADAPTOR AP-1 47 KD PROTEIN) (HA1 47 KD SUBUNIT) (CLATHRIN ASSEMBLY PROTEIN ASSEMBLY PROTEIN COMPLEX 1 MEDIUM CHAIN)	Contains protein domain (PF00928) - Adaptor complexes medium subunit family		glycoprotein	284488, 22278996, 284259, 35698052, 264905, 264906, 264907, 264908, 264909, 264510, 284512, 285008, 285009, 264910, 264591, 264592, 284593, 284594, 264758, 265019, 264760, 264681, 18108351, 264683, 264764, 18108354, 284766, 264768, 264769, 264689, 21908766, 21908767, 21908769, 29148629, 35695917, 265020, 265022, 33657023, 33657109, 18108370, 264628, 284628, 284631, 284632, 284635, 56182323, 60170394, 18108385, 264563, 284564, 264566, 284567
2456	88166700 (4911, 4912)	Novel Protein sim. GBank gil2588630 (AC003079) - Ankyrin- like; 54% similar to 2022340A (NID:g1092123) in exons spanning 43974 to 11551 of clone. [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat		kinase	264683

2457	94118375 (4913, 4914)	Novel Protein sim. GBank gij3025447 (AC004528) - R32184_3 [Homo sapiens]		UNCLASSIFIED	55181686, 264805, 264907, 264511, 264596, 55811386, 264682, 264684, 264685, 264687, 264691, 33657023, 264693, 35695855, 264636, 264555, 56182323, 264558, 56528488, 264583
2458	85675304 (4915, 4916)	Novel Protein sim. GBank gij2384942 (AF022985) - Similar to collagen [Caenorhabditis elegans]		UNCLASSIFIED	264691, 264693, 264634, 264559
2459	87551913 (4917, 4918)	Novel Protein sim. GBank gij5441942gbjAAD43187.1AC00499 - (AC004997) supported by mouse EST AA538043 (NID:g2284036) [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	transport	27485265
2460	94315289 (4919, 4920)	Novel Protein sim. GBank gij4929701gbjAAD34111.1AF15187 - (AF151874) CGI:116 protein [Homo sapiens]		kinase	65274572, 35696286, 22278996, 22278997, 60432049, 29331822, 29331824, 29331825, 66714117, 60432289, 29331826, 29331827, 29331828, 33656970, 29146499, 264102, 264109, 60433438, 265017, 265018, 265019, 264288, 21908785, 21908786, 21908789, 35695917, 265020, 264691, 33657023, 27485261, 18108374, 35695855, 87168518, 60432113
2461	87645147 (4921, 4922)	Novel Protein sim. GBank gij4426962gbjAAD20633 - (AF126052) Ar-like 2 binding protein BART1 [Homo sapiens]		UNCLASSIFIED	264259, 29331828, 264910, 18108351, 18108370, 18108374
2462	86988002 (4923, 4924)	Novel Protein sim. GBank gij5420387embjCAB46679.1 - (AJ243459) proteophosphoglycan [Leishmania major]			264909, 264758, 264684, 18108374, 264637, 18108385
2463	84388543 (4925, 4926)	Novel Protein sim. GBank gij5052516gbjAAD38588.1AF14561 - (AF145613) BcDNA GH03108 [Drosophila melanogaster]	Contains protein domain (PF00096) - Zinc finger, C2HZ type	UNCLASSIFIED	264681, 264566
2464	97219957 (4927, 4928)	Novel Protein sim. GBank gij5410300gbjAAD43021.1 - (AF100757) COP9 complex subunit 4 [Homo sapiens]	Contains protein domain (PF01399) - PCI domain	protease	264499, 52646842, 22278995, 35696286, 22278996, 22278997, 22278999, 264259, 29331822, 29331824, 66714117, 29331825, 29331826, 29331827, 35696052, 29331828, 264508, 52644045, 264510, 264511, 264512, 265008, 60170831, 264593, 52646317, 33109954, 33657084, 265017, 265018, 265019, 264762, 264448, 264764, 264288, 264766, 21905765, 21905766, 21905767, 21905768, 21906769, 265021, 33657023, 33657109, 18108370, 18108381, 60170394, 18108385, 22278902, 264486

2465	95357483 (4828, 4930)	Novel Protein sim. GBank gij4508401[ref]NP_002871.1 pRAF1 - v-raf-1 murine leukemia viral oncogene homolog 1	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	oncogene	18108392, 52644507, 52845156, 52866365, 22278994, 22278995, 35986286, 22278996, 22278998, 264259, 29331822, 29331824, 29331825, 60424269, 60432289, 29331827, 35986052, 29331828, 264907, 29331830, 52644045, 264909, 56182435, 264511, 265007, 265008, 265009, 264910, 33657402, 60433438, 55812038, 21908754, 33109954, 265010, 265011, 87168559, 264600, 265017, 265018, 265019, 18108351, 264369, 264288, 264695, 264767, 21908765, 21908767, 21908768, 55811957, 35695917, 265020, 265021, 60170815, 52644150, 33657023, 52645129, 33657109, 27486261, 27486264, 35695783, 264628, 263972, 18108374, 35695855, 264636, 264837, 60170394, 55256485, 87168518, 60432113, 264563, 264584, 264586, 264487 264369
2466	85681386 (4931, 4932)	Novel Protein sim. GBank gij4321619 gb AAD15788.1  - (AF051098) seven transmembrane domain orphan receptor [Mus musculus]			
2467	88059465 (4933, 4934)	Novel Protein sim. GBank gij3513300 (AC005595) - F16601_1, partial CDS [Homo sapiens]		UNCLASSIFIED	56994075, 264808, 21908768, 33657023
2468	87614696 (4935, 4936)	Novel Protein sim. GBank gij214345 pir j158106 - gene DMR-N9 protein - mouse (fragment)	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	29331824, 52844045, 265008, 264910, 265019, 21908765, 21908769, 265021 264288, 264628
2469	86294397 (4937, 4938)	Novel Protein sim. GBank gij5420389 emb CAB46680.1  - (AJ243460) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264592, 264692, 264555, 264556, 264557, 264558, 264559, 18108385, 264482
2470	80223831 (4939, 4940)	Novel Protein sim. GBank gij5419882 emb CAB46424.1  - (AL096749) DKFZp434G153 [Homo sapiens]		UNCLASSIFIED	65274572, 35986286, 29331827, 265007, 264592, 33109854, 265018, 265019, 264288, 55811957, 265020, 264693, 55811576, 56182323
2472	85060811 (4943, 4944)	Novel Protein sim. GBank gij4929747 gb AAD34134.1 AF15189 - (AF151897) CGI-139 protein [Homo sapiens]		UNCLASSIFIED	264092, 60432049, 29331825, 60433356, 265010, 265011, 18108351, 264764, 264288, 264692, 65274620, 18108370, 18108372, 18108374, 264634, 18108385
2473	95421509 (4945, 4946)	Novel Protein sim. GBank gij4538009 emb CAB39630.1  - (AL049481) putative protein [Arabidopsis thaliana]			60424179, 65274572, 22278998, 60424289, 29331826, 265008, 60433356, 60433438, 265010, 18108351, 264448, 264288, 264687, 264689, 265021, 264692, 65274620, 60431528, 65274791, 264556, 56182323, 60432113
2474	94315616 (4947, 4948)	Novel Protein sim. GBank gij3252827 (AC004382) - Unknown gene product [Homo sapiens]			65274572, 56994075, 264259, 28331828, 60170831, 265017, 265018, 265019, 284683, 264369, 265020, 264693, 264563, 264564

2475	94321693 (4949, 4950)	Novel Protein sim. GBank gi 1216485 (U48852) - HT protein [Cricetulus griseus]	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	264259, 29331822, 265008, 265007, 265010, 265011, 264448, 264288, 264369, 264685, 264686, 18108357, 264768, 18108362, 264693, 18108370, 18108374, 18108379, 35696423, 83373044, 18108383, 18108385, 264564, 264565, 264567
2476	94315618 (4951, 4952)	Novel Protein sim. GBank gi 3252827 (AC004382) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	264259, 60424289, 66714117, 264905, 265006, 264511, 265008, 265009, 264758, 265010, 265011, 18108351, 264681, 264369, 264288, 264689, 21906767, 265020, 18108374, 264639, 18108382, 83373044, 18108385, 87168518
2477	20718974 (4953, 4954)			UNCLASSIFIED	263978
2478	17659165 (4955, 4956)			UNCLASSIFIED	265017
2479	94314569 (4957, 4958)	Novel Protein sim. GBank gi 164232[dbj]BAA110821 - (D67066) N-WASP [Bos taurus]	im7	UNCLASSIFIED	56894075, 22278998, 21906754, 264682, 21906785
2480	95295605 (4959, 4960)			UNCLASSIFIED	264905, 264907, 264765
2481	94718481 (4961, 4962)	Novel Protein sim. GBank gi 5689469[dbj]BAA83018.1 - (AB028989) KIAA1088 protein [Homo sapiens]	collagen	UNCLASSIFIED	55274572, 56182575, 22278997, 264094, 264259, 29331822, 29331824, 66714117, 28331827, 35696052, 264508, 264805, 264908, 264907, 264908, 52644045, 264909, 56182435, 265008, 264910, 33657402, 55812038, 264758, 265010, 265011, 265017, 55811957, 265020, 264691, 264692, 264693, 264628, 55811576, 264630, 264634, 264635, 264636, 264637, 264556, 264558, 56182323, 83373044, 60432113, 22278002
2482	87393165 (4963, 4964)	Novel Protein sim. GBank gi 321249[pir]S28407 - guanine nucleotide-exchange activator CDC25 homolog - mouse		UNCLASSIFIED	29331822, 29331824, 29331825, 29331827, 264508, 264905, 264509, 264906, 264907, 264908, 264511, 264591, 264768, 264693, 264631, 264632, 264636, 264638, 264639, 264583
2483	87731583 (4965, 4966)			UNCLASSIFIED	264488, 22278995, 264093, 264095, 60432049, 60433356, 60433438, 284448, 264288, 263967, 18108370, 18108385, 18108388, 264482
2484	94187774 (4967, 4968)	Novel Protein sim. GBank gi 728831[sp]P39188[ALU1]_HUMAN - !!! ALU SUBFAMILY J WARNING ENTRY !!!	kinase	UNCLASSIFIED	264553
2485	87786556 (4969, 4970)	Novel Protein sim. GBank gi 1185397 (U25281) - SH3 domain binding protein [Rattus norvegicus]		UNCLASSIFIED	22278995, 22278996, 22278997, 22278999, 264259, 60432049, 29331824, 60432289, 29331827, 285007, 264910, 264593, 264600, 264603, 264604, 265019, 264448, 264288, 264685, 264686, 264769, 264689, 35695917, 265022, 264692, 264693, 56182323

2486	87748978 (4971, 4972)	Novel Protein sim. GBank gii2662167[dbj BA023715] - (AB007803) KIAA0443 [Homo sapiens]			265017, 264555
2487	95343105 (4973, 4974)	Novel Protein sim. GBank gii46455[sp P35287 RB14_RAT - RAS-RELATED PROTEIN RAB-14]	Contains protein domain (PF00071) - Ras family	glycoprotein	22278996, 35696286, 22278997, 22278998, 22278999, 264092, 264259, 28331822, 35696052, 284108, 284905, 284907, 29331830, 264908, 265008, 264511, 265008, 265009, 60433438, 21906754, 33109954, 87188558, 285018, 284881, 284288, 284887, 21906765, 21906766, 21906767, 21906788, 21906769, 35695917, 265021, 265022, 264534, 33657023, 284882, 33657109, 263972, 18108377, 35698423, 35698855, 60170394, 18108385, 56526488, 22279000, 22279002, 264563, 264482, 264565, 20281189, 18108391
2488	87652451 (4975, 4976)			UNCLASSIFIED	264910, 264448, 264288, 264684, 264691, 264634
2489	82990585 (4877, 4978)	Novel Protein sim. GBank gii4886439[emb CAB43355.1] - (AL050253) hypothetical protein [Homo sapiens]	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	complementrecept	284688, 284693, 55811576, 22279002
2490	88068609 (4979, 4980)	Novel Protein sim. GBank gii2588624 (AC003083) - Rap2 interacting protein-like: similar to U73941 (PID:g1916018) [Homo sapiens]		UNCLASSIFIED	284907, 265008, 22279002
2491	91242116 (4981, 4982)	Novel Protein sim. GBank gii728832[sp P39189 ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII]		tm7	264259, 29331826, 265008, 264762, 18108370, 18108376, 18108379
2492	95308202 (4983, 4984)	Novel Protein sim. GBank gii3355303 (AF001549) - Unknown gene product [Homo sapiens]		transcriptfactor	284488, 22278998, 22278999, 29331828, 264591, 33109954, 265017, 55811150, 21906764, 21906788, 264682, 60431528, 87188558, 60432113, 22279000

2483	95422415 (4985, 4986)	Novel Protein sim. GBank gi 4240307 cbj BAA74932.1  - (AB020716) KIA0909 protein [Homo sapiens]	Contains protein domain (PF01424) - Struct R3H domain	18108394, 264887, 65274572, 56182575, 22278995, 56984075, 60432048, 29331822, 29331824, 29331825, 29331826, 29331827, 29146498, 264508, 264905, 264509, 264906, 264907, 29331830, 264909, 264510, 265006, 264511, 265007, 264512, 265008, 265009, 264910, 21808754, 265011, 284600, 265017, 265018, 264604, 264605, 265019, 55811150, 264762, 18108351, 264681, 264448, 264683, 264369, 264288, 18108355, 18108357, 264687, 21806765, 21906766, 21906767, 21906768, 21906769, 265020, 264691, 264692, 33657023, 33657348, 18108370, 18108374, 18108376, 55810764, 18108379, 65274791, 264630, 264632, 264634, 264635, 264636, 264555, 264637, 264557, 264558, 264639, 264559, 83373044, 18108385, 87168518, 60432113, 22279000, 22279002, 264482, 264586, 264488
2494	30793119 (4987, 4988)	Novel Protein sim. GBank gi 5420389 emb CA846660.1  - (AJ243460) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	264907, 264601
2495	94234551 (4989, 4990)	Novel Protein sim. GBank gi 5420389 emb CA846660.1  - (AJ243460) proteophosphoglycan [Leishmania major]	collagen	263994, 22278997, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 265006, 265009, 264595, 264604, 264448, 264682, 264764, 264288, 264685, 264768, 264769, 264689, 265020, 264692, 65274820, 264629, 55810764, 35696423, 55811576, 264636, 264637, 18108385, 22278000, 264584, 264587, 264488
2496	80018763 (4991, 4992)	Novel Protein sim. GBank gi 4808220 emb CA842832.1  - (AL022315) dJ11715.1 (PUTATIVE novel protein) [Homo sapiens]	struct	29147620, 264805, 265006, 265007, 18108348, 18108362, 18108370, 18108374, 264555, 264556, 18108381, 18108383, 18108388
2497	61723554 (4993, 4994)		UNCLASSIFIED	52644507, 22278996, 22278999, 29331824, 29331828, 33657402, 21908754, 87168474, 265019, 264389, 264689, 21908785, 21908768, 21906767, 21906768, 265020, 33657023, 18108376, 18108387
2498	87724633 (4995, 4996)	Novel Protein sim. GBank gi 1200503 (U47924) - B [Homo sapiens]	UNCLASSIFIED	29331827, 264512, 264910, 264288, 18108374, 35694855
2499	94665125 (4997, 4998)	Novel Protein sim. GBank gi 3510234 (AC005581) - R31237.1, partial CDS [Homo sapiens]	Contains protein domain (PF00059) - kinase Eukaryotic protein kinase domain	264909, 55812038, 264631, 264637, 264558

2500	94648324 (4989, 5000)	Novel Protein sim. GBank gi 3881275 emb CAA21725  - (AL032655) predicted using GeneFinder; similar to Inositol monophosphatase family; cDNA EST K255e11.5 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00459) - Inositol monophosphatase family	transport	52844507, 52845156, 22278995, 56994075, 35892866, 22278998, 284259, 52845080, 29331824, 29331825, 68714117, 60432289, 29331826, 29331827, 35696052, 29331828, 284508, 264509, 264510, 264512, 33657402, 60433438, 21906754, 52644296, 87168474, 87168559, 264603, 264681, 284448, 264683, 264288, 264389, 52644229, 264689, 21906765, 21906786, 21906787, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 52644150, 33857023, 264693, 33657182, 35695783, 35696423, 35695855, 52644332, 83373044, 18108387, 87168818, 22279002
2501	94303896 (5001, 5002)	Novel Protein sim. GBank gi 4928615 gb A034058.1 AF15183 - (AF151831) CGI-73 protein [Homo sapiens]	Contains protein domain (PF00051) - BTB/POZ domain	dna_mn_bind	65274572, 56182575, 35696286, 22278996, 56994075, 22278997, 60432049, 264259, 29331822, 29331824, 29331826, 29331827, 35696052, 264905, 284908, 264907, 264908, 264909, 56182435, 264510, 264511, 265007, 264910, 264591, 60432229, 33657402, 60433396, 264595, 55812038, 264758, 264596, 87168474, 87168559, 264600, 264601, 264602, 265017, 264604, 265018, 264605, 265019, 18108351, 264448, 264389, 264288, 264766, 18108357, 21906765, 21906766, 21906767, 21906768, 29148629, 35695917, 264692, 33657023, 264629, 35695423, 55811576, 35695855, 264630, 264634, 264635, 264555, 264636, 264638, 264558, 60170394, 83373044, 18108385, 18108387, 87168518, 60432113, 22279002, 264568
2502	90993716 (5003, 5004)	Novel Protein sim. GBank gi 3041847 (AC004542) - OXYSTEROL-BINDING PROTEIN-like; similar to P22059 (PID.g129308) [Homo sapiens]	Contains protein domain (PF01237) - Oxysterol-binding protein	UNCLASSIFIED	65274572, 264907, 56182435, 285007, 264592, 264760, 18108351, 264448, 284369, 264288, 264684, 264686, 55811957, 265021, 264692, 33657109, 263973, 55811576, 264635, 264553, 264556, 264557, 264558, 55812232, 264559, 87168518, 264563, 264482
2503	87878345 (5005, 5006)	Novel Protein sim. GBank gi 2196874 emb CAA72638  - (Y11596) BRX protein [Mus musculus]			264905, 264907, 264512, 265008, 265011, 18108351, 264448, 264288, 29148627, 284693, 18108370, 18108374, 18108385



2504	87868706 (5007, 5008)	Novel Protein sim. GBank gii550420jembjCAA48220j - (X68101) Irq (Rattus norvegicus)		264488, 52644507, 52645156, 52646842, 22278994, 264259, 52845080, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 264908, 264908, 52644045, 265009, 60433355, 33657402, 60433438, 264595, 33109954, 87168474, 265017, 265018, 264448, 264288, 264768, 52644228, 21906765, 21906768, 21906767, 21906768, 52844150, 264892, 27486281, 27486282, 27486284, 27486285, 35695763, 35696423, 35695855, 52644332, 56182323, 18108387, 87168518, 60432113, 22279002, 264564	UNCLASSIFIED	264605
2505	8760559 (5009, 5010)			264488, 52644507, 52645156, 52646842, 52647452, 22278995, 56994075, 22278996, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 29331826, 29331828, 264508, 56182435, 264112, 264593, 60433356, 55812038, 21906754, 265011, 265017, 265018, 265019, 284605, 264762, 18108351, 264448, 264288, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 35695817, 265020, 265021, 265022, 60170615, 33657023, 27486284, 18108378, 35695855, 264637, 83373044, 18108385, 87168518, 60432113, 22279000, 264563, 264482, 264565	UNCLASSIFIED	264605
2506	91232328 (5011, 5012)	Novel Protein sim. GBank gii2137562pjiI49635 - mouse Dhml protein - mouse		264488, 52644507, 52645156, 52646842, 52647452, 22278995, 56994075, 22278996, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 29331826, 29331828, 264508, 56182435, 264112, 264593, 60433356, 55812038, 21906754, 265011, 265017, 265018, 265019, 284605, 264762, 18108351, 264448, 264288, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 35695817, 265020, 265021, 265022, 60170615, 33657023, 27486284, 18108378, 35695855, 264637, 83373044, 18108385, 87168518, 60432113, 22279000, 264563, 264482, 264565	UNCLASSIFIED	264605
2507	95316233 (5013, 5014)	Novel Protein sim. GBank gii5174489jleNP_006033.1pKIAA - histone deacetylase 6	Contains protein domain (PF00850) - Histone deacetylase family	264488, 52644507, 52645156, 52646842, 52647452, 22278995, 56994075, 22278996, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 29331826, 29331828, 264508, 56182435, 264112, 264593, 60433356, 55812038, 21906754, 265011, 265017, 265018, 265019, 284605, 264762, 18108351, 264448, 264288, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 35695817, 265020, 265021, 265022, 60170615, 33657023, 27486284, 18108378, 35695855, 264637, 83373044, 18108385, 87168518, 60432113, 22279000, 264563, 264482, 264565	UNCLASSIFIED	264605
2508	95315505 (5015, 5016)	Novel Protein sim. GBank gii4826433jembjCAB42889.1 - (AL031447) dJ126A5.2.1 (novel protein) (Isolform 1) (Homo sapiens)		264488, 52644507, 52645156, 52646842, 52647452, 22278995, 56994075, 22278996, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 29331826, 29331828, 264508, 56182435, 264112, 264593, 60433356, 55812038, 21906754, 265011, 265017, 265018, 265019, 284605, 264762, 18108351, 264448, 264288, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 35695817, 265020, 265021, 265022, 60170615, 33657023, 27486284, 18108378, 35695855, 264637, 83373044, 18108385, 87168518, 60432113, 22279000, 264563, 264482, 264565	UNCLASSIFIED	264605
2509	87813741 (5017, 5018)	Novel Protein sim. GBank gii1263289 (U47655) - fibroin-4 (Araneus diadematus)		264488, 52644507, 52645156, 52646842, 52647452, 22278995, 56994075, 22278996, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 29331826, 29331828, 264508, 56182435, 264112, 264593, 60433356, 55812038, 21906754, 265011, 265017, 265018, 265019, 284605, 264762, 18108351, 264448, 264288, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 35695817, 265020, 265021, 265022, 60170615, 33657023, 27486284, 18108378, 35695855, 264637, 83373044, 18108385, 87168518, 60432113, 22279000, 264563, 264482, 264565	UNCLASSIFIED	264605

2510	95421378 (5018, 5020)	Novel Protein sim. GBank gij3293537[gb AAC25782.1] - (AF071059) zinc finger RNA binding protein [Mus musculus]		gna_ma_bind	65274573, 22278994, 22278996, 22278998, 22278999, 60432048, 264259, 29331822, 29331824, 29331825, 66714117, 60432289, 29331826, 33656970, 264908, 66712502, 265007, 264910, 60170831, 60432229, 60433396, 60433438, 21908754, 87168474, 265017, 265018, 264448, 264288, 21806767, 21906768, 21906769, 55811957, 35695917, 265020, 265022, 264691, 33657023, 264693, 65274620, 33657109, 33657182, 27486262, 33657349, 18108370, 35695855, 264555, 56182323, 83373044, 60432113, 22279002, 265017, 21908784, 265020, 264692
2511	87384281 (5021, 5022)	Novel Protein sim. GBank gij4323152[gb AAD16228.1] - (AF098863) Ets-protein Spi-C [Mus musculus]			
2512	88084771 (5023, 5024)	Novel Protein sim. GBank gij4502075[tr NP_001135.1]pAMFR - autocrine motility factor receptor	Contains protein domain (PF000097) - Zinc finger, C3HC4 type (RING finger)	transport	22278999, 264259, 29331825, 29331826, 29146499, 264907, 264909, 265006, 265008, 264591, 60432229, 21906754, 264763, 264683, 264768, 18108357, 264689, 21906769, 264693, 18108370, 263972, 18106374, 264558, 22279000
2513	95357843 (5025, 5026)	Novel Protein sim. GBank gij3004657 (AF017777) - bobby sox [Drosophila melanogaster]		UNCLASSIFIED	60424178, 52845166, 18108394, 22278894, 35686286, 56984075, 22278898, 29331822, 29331824, 60424269, 29331825, 29331827, 33656970, 60431735, 33657084, 87168559, 265017, 264448, 264369, 56181562, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 33657023, 18108398, 33657109, 27486281, 27486282, 33657349, 18108374, 55810784, 35698423, 58182323, 264558, 18108385, 264510
2514	88094578 (5027, 5028)	Novel Protein sim. GBank gij2258437 (AF008197) - syncollin [Rattus norvegicus]		UNCLASSIFIED	
2515	87994509 (5029, 5030)	Novel Protein sim. GBank gij375727[emb CAA16783] - (AL022727) dJ8019.7 (olfactory receptor-like protein [hs6M1-3]) [Homo sapiens]	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	- [m7]	
2516	87766808 (5031, 5032)			UNCLASSIFIED	264259, 29148498, 264905, 264288, 29148629, 35695917, 27488281, 264634, 264091, 29331824, 29331825, 29331826, 29331828, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 264910, 33657402, 264757, 33109954, 265017, 265018, 264605, 264760, 264762, 264763, 264766, 264768, 264769, 33657109, 33657182, 264828, 55811576, 35698423, 264631, 264634, 264637, 264638, 264639, 87168518, 22279002, 264564
2517	87764968 (5033, 5034)	Novel Protein sim. GBank gij4220527[emb CAA23000] - (AL035356) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	

2518	94147410 (5035, 5036)	Novel Protein sim. GBank gi 4928991 gb AAO34056.1 JAF15181 - (AF15181) CGI-61 protein [Homo sapiens]	Contains protein domain (PF00018) - SH3 domain	UNCLASSIFIED	35696286, 264259, 29331822, 29331824, 29331825, 60432289, 29331828, 29331827, 35896052, 29331828, 264907, 264909, 284511, 265007, 60432229, 60433356, 60433438, 55812038, 265010, 265017, 264448, 264288, 264689, 21906768, 21906769, 265022, 52644150, 264689, 18108370, 263972, 264555, 56182323, 93373044, 18108385, 60432113, 264088 264259, 66714117, 29331826, 29331827, 29331828, 264907, 66712502, 265006, 265008, 264594, 265010, 265011, 265018, 264288, 21906769, 265020, 60431528, 55811576, 65274791, 264632, 264555, 264536, 22279002, 264564 264259, 264908, 264910, 264682, 21906769, 265020, 264563
2519	94326180 (5037, 5038)	Novel Protein sim. GBank gi 4263748 gb AAD15420 - (AC004883) similar to KIAA0766; similar to P1Dg3882233 [Homo sapiens]	kinase		
2520	87413235 (5039, 5040)	Novel Protein sim. GBank gi 4826722 ref NP_005085.1 pFATP - fatty acid transport protein 4	transport		
2521	95318244 (5041, 5042)	Novel Protein sim. GBank gi 5174489 ref NP_006035.1 pKIAA - histone deacetylase 6	Contains protein domain (PF00850) - Histone deacetylase family	histone	264488, 264489, 263994, 65274572, 22278995, 22278998, 264259, 29331822, 29331826, 264508, 264905, 264509, 264806, 264907, 66712502, 264511, 265008, 265007, 264591, 264592, 264593, 264594, 264595, 264596, 264681, 264448, 264763, 264682, 264764, 264684, 264369, 264288, 264685, 264686, 21906768, 55811957, 264692, 264693, 27486251, 18108370, 264628, 264629, 18108374, 55811576, 35696423, 35695855, 264632, 264558, 18108385, 65274727, 60432113, 264563, 264564, 264565, 264568, 264567 264489, 22278997, 20281171, 21906754, 35695917, 263987, 263976, 263981, 20281169 263969
2522	87754052 (5043, 5044)	Novel Protein sim. GBank gi 458001 gb AAD24201.1 U81002_ - (U81002) TRAF4 associated factor 1 [Homo sapiens]	transport		
2523	95340467 (5045, 5046)	Novel Protein sim. GBank gi 1809327 (U76374) - skm- BOP2 [Mus musculus]	Contains protein domain (PF01753) - MYND finger		56994075, 22278998, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331828, 29331828, 29331830, 56182435, 264512, 265008, 60170831, 33657402, 265010, 87168559, 265019, 264288, 21906765, 21906769, 35695917, 265020, 265021, 265022, 52644150, 264691, 33657023, 33657108, 27486281, 35696423, 65274791, 264559, 83373044, 56528486, 87168518, 264567
2524	95340469 (5047, 5048)				

2525	94126928 (5049, 5050)	Novel Protein sim. GBank gjl2073564 (U00223) - eukaryotic initiation factor eIF-2 alpha kinase; DGCN2 [Drosophila melanogaster]		kinase	264488, 2227897, 2227898, 60432049, 60432289, 29331828, 264905, 265008, 55812038, 21905754, 265019, 264369, 21908765, 21908766, 21908767, 21908769, 35695917, 285020, 265021, 33657109, 60431528, 83373044, 60432113, 22278000, 22279002, 284585
2526	95289404 (5051, 5052)	Novel Protein sim. GBank gjl4586528(jb)BAA76836.1] - (AB023209) KIA0992 protein [Homo sapiens]	Contains protein domain (PF00236) - Ribosomal protein L14	ribosomal prot	60424178, 284768, 264687, 264769, 264689, 65274572, 21908767, 56182575, 21908768, 21908769, 55811657, 22278994, 22278995, 35695286, 35695917, 22278996, 22278997, 265020, 22278998, 265021, 22278999, 265022, 264690, 264691, 60432048, 264259, 264097, 33657023, 29331822, 29331824, 60432289, 29331826, 29331827, 29331828, 27486282, 264508, 284509, 284805, 264907, 18108370, 66712502, 60431528, 264828, 264909, 18108372, 18108374, 56182435, 18108376, 55810764, 55811576, 35696423, 35695855, 265006, 265007, 264512, 265008, 265009, 264634, 264635, 60431850, 284636, 264555, 264592, 60431735, 264638, 33657402, 56182323, 60433356, 60433438, 264595, 55812038, 264596, 264758, 83373044, 52646317, 18108385, 33657084, 18108387, 55811386, 55274727, 56528488, 87168518, 60432113, 265017, 22279000, 265018, 265019, 264584, 18108351, 264448, 284586, 284288, 264486, 264587, 264766, 264510, 264512, 264630, 264591, 264592, 264259, 264594, 264595, 264603, 264605, 18108351, 264585, 264369, 18108354, 55812038, 265017, 264689, 35695917, 35695763, 60431528, 60432113, 22278002
2527	88094580 (5053, 5054)	Novel Protein sim. GBank gjl2258437 (AF008197) - syncollin [Rattus norvegicus]		UNCLASSIFIED	
2528	88078380 (5055, 5056)	Novel Protein sim. GBank gjl2085786 (AC002066) - similar to zinc finger 5 protein from Gallus gallus, U51640 (P1Dg1399185) [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	dna_rna_bind	
2529	86670926 (5057, 5058)	Novel Protein sim. GBank gjl3786433 (AF098505) - similar to Arabidopsis thaliana male sterility protein 2 (SW:Q08891) [Caenorhabditis elegans]		synthase	
2530	80259878 (5059, 5060)				264369, 264556
2531	87768931 (5061, 5062)			UNCLASSIFIED	29331822, 29331824, 60432289, 264508, 264509, 264908, 265011, 264769, 21908768, 33657023, 87168518, 22279000
2532	87419778 (5063, 5064)	Novel Protein sim. GBank gjl2884625(jemb)CAA189721 - (AL021811) putative protein [Arabidopsis thaliana]			264583
2533	87000255 (5065, 5066)	Novel Protein sim. GBank gjl437181 (U02289) - GTPase-activating protein [Caenorhabditis elegans]		UNCLASSIFIED	264555

2534	87332322 (5067, 5068)	Novel Protein sim. GBank gij3452473 (AF084205) - serine/threonine protein kinase TAO1 (Rattus norvegicus)		UNCLASSIFIED	264258, 35886052, 264905, 265017, 21906789, 265020, 265022, 33657109, 22279000
2535	91225056 (5069, 5070)	Novel Protein sim. GBank gij446831 (emb CA837932  - (AL03432) dJ468N24.2.1 (PUTATIVE novel protein) (isolom 1) [Homo sapiens])			65274572, 35898286, 60432289, 29331828, 66712502, 265006, 60432229, 265017, 265018, 265019, 264288, 264359, 264689, 21906788, 265020, 265021, 264636, 60170394, 22279002
2536	94218540 (5071, 5072)	Novel Protein sim. GBank gij28836 (sp P39193 ALU6_HUMAN - IIII ALU SUBFAMILY SP WARNING ENTRY IIII)		kinase	18108398, 56182575, 3566286, 22278997, 22278989, 60432049, 264259, 29331824, 29331826, 29331827, 28331828, 264905, 264511, 265009, 264910, 264598, 52646317, 18108351, 264681, 264683, 18108354, 264288, 264687, 264786, 264689, 21906785, 21906788, 21906787, 265021, 52645129, 33657109, 18108374, 18108380, 56182323, 18108381, 18108388, 87168518, 60432113, 22279000, 22279002, 264567, 18108391
2537	95422283 (5073, 5074)	Novel Protein sim. GBank gij4557026 (ref NP_003913.1 pHERC - guanine nucleotide exchange factor p532)		ubiquitin	65274572, 35662866, 29331822, 29331825, 29331827, 29331828, 35696052, 264906, 66712502, 264909, 265008, 265011, 264760, 264288, 264685, 35695917, 60170615, 264691, 33657023, 65274820, 33657109, 18108374, 35696423, 35695855, 264636, 264558, 60170394, 56182323, 83373044
2538	36853454 (5075, 5076)			UNCLASSIFIED	
2539	94144916 (5077, 5078)			UNCLASSIFIED	22278996, 22278999, 29331822, 29331825, 29331828, 29146499, 264908, 264112, 60170831, 87168559, 264604, 265019, 264685, 264788, 87168518, 22279000, 264565, 264568
2540	94218545 (5079, 5080)	Novel Protein sim. GBank gij1362647 (piti S53876 - sex-regulated protein janus A - fruit fly (Drosophila pseudoobscura))		UNCLASSIFIED	22278997, 29331828, 265008, 265009, 264756, 265010, 18108351, 264683, 264288, 21906785, 35695917, 265020, 18108374, 264567
2541	95308238 (5081, 5082)	Novel Protein sim. GBank gij171168 (sp P54797 T10_MOUSE - SER/THR-RICH PROTEIN T10 IN DCCR REGION)		UNCLASSIFIED	264488, 264788, 264689, 264511, 20281171, 264634, 264635, 264691, 264639, 29331824, 264603, 264604, 264905, 264907, 264908, 264788

2542	85288162 (5083, 5084)	Novel Protein sim. GBank gij5225320igbIAAD0850.1(AF08310 - (AF083107) siruin type 2 [Homo sapiens])	Contains protein domain (PF00220) - Neurohypophysis hormones, N- terminal Domain	UNCLASSIFIED	264488, 18108394, 52646385, 52646842, 85274572, 22278994, 35898286, 22278996, 264259, 52645080, 29331822, 29331824, 29331827, 35898052, 33656970, 264907, 264809, 52644045, 264510, 265008, 264512, 265007, 265008, 265009, 264910, 60431735, 52646317, 52644288, 265010, 265011, 265018, 265019, 18108351, 264683, 264288, 264685, 264687, 52644228, 264769, 21908788, 21908787, 21908789, 52644150, 33657023, 33657108, 52645129, 33657182, 27486261, 27486264, 33657349, 35895783, 18108374, 35896423, 35895855, 264831, 264634, 264635, 264558, 83373044, 18108385, 18108387, 87168518, 264583, 264584
2543	94139088 (5085, 5086)	Novel Protein sim. GBank gij5418857[emb/CAB46374.1] - (AL096723) hypothetical protein [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)		65274572, 56182575, 22278999, 264259, 29331826, 264907, 264510, 264511, 264592, 264595, 264784, 264389, 264288, 264894, 264766, 264689, 21906765, 21906767, 21906769, 60170515, 264692, 264693, 55811578, 65274791, 264636, 264558, 18108381, 60170394, 264639, 18108385, 60432113, 22279000
2544	84218549 (5087, 5088)	Novel Protein sim. GBank gij2498110spIQ63191JAEIP_RAT - APICAL ENDOSOMAL GLYCOPROTEIN PRECURSOR	Contains protein domain (PF00629) - MAM domain.	glycoprotein	18108397, 52846385, 22278997, 264259, 60432049, 29331822, 29331825, 29331826, 29331827, 29331828, 264905, 264908, 265006, 265007, 265008, 87168559, 265017, 265018, 265019, 18108351, 264448, 264686, 264687, 264688, 21906765, 265020, 265021, 18108370, 18108374, 18108376, 18108381, 18108385, 18108387, 56526486, 22279000, 264482, 264563, 264567
2545	87742645 (5089, 5090)	Novel Protein sim. GBank gij3327046[dbj BA31591] - (AB014516) KIAA0516 protein [Homo sapiens]			29331825, 264906, 265009, 60170831, 265017, 264389, 21908767, 60170815, 264682, 33657109
2546	86083861 (5091, 5092)	Novel Protein sim. GBank gij2996032 (AF054586) - brain finger protein [Rattus norvegicus]	Contains protein domain (PF00087) - Zinc finger, C2HC4 type (RING finger)	UNCLASSIFIED	29331824, 265007, 22279002

2547	94136893 (5093, 5094)	Novel Protein sim. GBank gjl4928607/gbJAAD34064.1 AF15182 - (AF151827) CGI-69 protein (Homo sapiens)	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	264488, 18108394, 52846842, 18108397, 56182575, 22278995, 56994075, 22278996, 22278997, 22278998, 264259, 29331822, 29331824, 29331826, 60432289, 29331827, 35696032, 29331828, 264104, 264508, 264905, 264906, 264908, 66712502, 264909, 56182435, 285006, 285007, 284512, 285008, 285009, 60170831, 60432228, 60431735, 284594, 60433438, 21906754, 52846317, 265010, 265011, 264600, 264601, 265018, 265019, 264760, 18108351, 264692, 264448, 264288, 284369, 264684, 264686, 264687, 56181562, 284688, 284689, 21906765, 21906766, 21906767, 21906768, 29148627, 21906769, 55811957, 265020, 265021, 265022, 284890, 264891, 18108382, 264692, 264693, 27488261, 18108370, 18108374, 55810764, 55811576, 35696423, 35695855, 264635, 264636, 264555, 264637, 263981, 264557, 18108380, 264638, 56182323, 284558, 284559, 83373044, 18108385, 87168518, 22278002, 264564, 264568, 264486
2548	88179079 (5095, 5096)			UNCLASSIFIED	264488, 18108394, 52846365, 22278994, 35686286, 56994075, 22278997, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 60432289, 28331827, 28331828, 56182435, 284511, 265007, 284512, 60433356, 87168559, 264684, 264369, 52844229, 265021, 33657023, 264692, 18108374, 52844332, 264557, 18108380, 18108381, 18108382, 18108384, 18108385, 60432113, 22279000, 22279002, 284563, 264567
2549	94136893 (5097, 5098)	Novel Protein sim. GBank gjl728637/islp39194 ALU7_HUMAN -      ALU SUBFAMILY SQ WARNING ENTRY	Contains protein domain (PF00412) - LIM domain containing proteins	struct	56182575, 22278998, 22278997, 22278998, 22278999, 264259, 284508, 284908, 29331830, 285009, 285010, 265018, 264688, 21906764, 21906765, 21906766, 21906767, 21906769, 265020, 265021, 52844150, 264691, 18108368, 60431602, 18108376, 35696423, 56182323, 18108387, 264587, 56182575, 29331822, 264105, 264512, 18108351, 35695917, 264637, 264638
2550	87778584 (5099, 5100)	Novel Protein sim. GBank gjl2143886/pil152523 - nucleoporin p62 homolog - rat (fragment)		UNCLASSIFIED	

2551	95308400 (5101, 5102)	Novel Protein sim. GBank gi 4337103 gb AAD18079  - (AF129758) NG26 [Homo sapiens]	Contains protein domain (PF00561) - UNCLASSIFIED alpha/beta hydrolase fold	18108396, 65274572, 22278995, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 264905, 56182435, 265007, 60433438, 55812038, 21808754, 65274444, 265017, 265018, 264805, 265019, 264288, 21908768, 21908788, 21908789, 265020, 60170615, 264693, 33657109, 35698423, 264638, 56182323, 83373044, 22279000
2552	95332620 (5103, 5104)		UNCLASSIFIED	56182575, 35696286, 29331824, 29331828, 35696032, 29331828, 264508, 264907, 56182435, 265008, 264591, 33109954, 264760, 55811957, 35695917, 33657023, 33657109, 18108374, 55811576, 35698423, 35695855, 56182323, 264558
2553	95308243 (5105, 5106)	Novel Protein sim. GBank gi 171185 sp P54797 T10 MOUSE - SER/THR-RICH PROTEIN T10 IN DSCR REGION	UNCLASSIFIED	264688, 264488, 263976, 264768, 29331828, 35696032, 35698423, 264601, 264511, 264602, 264910, 264634, 264760, 264555, 264762, 264908, 264592, 264691, 264566, 264808, 264684, 264567, 264908, 264768
2554	87761520 (5107, 5108)	Novel Protein sim. GBank gi 728835 sp P38182 ALUS_HUMAN - III ALU SUBFAMILY SC WARNING ENTRY IIII	cadherin	22278997, 29331822, 264508, 21808769, 33657023, 33657109, 56182323
2555	87627551 (5108, 5110)	Novel Protein sim. GBank gi 488431 gemb CAB43260.1  - (AL050084) hypothetical protein [Homo sapiens]	nuclease	29331824, 263972
2556	87645533 (5111, 5112)	Novel Protein sim. GBank gi 4106984 (AC003038) - R30923.1 [Homo sapiens]	Contains protein domain (PF00514) - UNCLASSIFIED Armadiol/beta-catenin-like repeats	22278998, 264508, 33657402, 264683, 264684, 264768, 264689, 33657023, 33657109, 35695855, 264558, 264567, 264595
2557	79437803 (5113, 5114)	Novel Protein sim. GBank gi 19110 sp P03211 EBN1_EBV - EBNA-1 NUCLEAR PROTEIN	UNCLASSIFIED	22278997, 29331824, 66714117, 29331825, 264906, 264511, 265018, 264448
2558	87617591 (5115, 5116)	Novel Protein sim. GBank gi 453899 gemb CAB39619.1  - (AL049481) AIG1-like protein [Arabidopsis thaliana]	UNCLASSIFIED	22278997, 29331822, 29331828, 60433356, 265011, 264288, 264755, 264766, 264769, 21908785, 21908786, 60432113, 264482
2559	88096382 (5117, 5118)	Novel Protein sim. GBank gi 5051399 gemb CAB44985.1  - (AL078630) 573K1.3 (nm17M1-4 (novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor LIKE protein))) [Mus musculus]	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	
2560	87994530 (5118, 5120)	Novel Protein sim. GBank gi 5326825 gb AAD42056.1 AF044953 NADH:ubiquinone oxidoreductase PGIV subunit [Homo sapiens]	UNCLASSIFIED	22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331826, 29331827, 29331828, 265007, 60432228, 87168558, 265017, 265018, 265019, 264689, 21908766, 21908768, 35695917, 265020, 33657023, 33657109, 18108374, 264634, 264559, 18108385, 87168518, 22279002



2562	87645539 (5123, 5124)	Novel Protein sim. GBank gij106984 (AC003038) - R30923_1 [Homo sapiens]		UNCLASSIFIED	56994075, 22278996, 22278997, 22278999, 264259, 29331822, 60432288, 33657402, 60433356, 21908765, 55811957, 60170815, 33657023, 264693, 35695855, 87168518, 264486, 35696286, 22278999, 264259, 29331822, 29331824, 35696052, 284508, 264907, 264908, 264909, 52644045, 284510, 264511, 265009, 264910, 264591, 264593, 33657402, 265017, 265018, 265019, 18108351, 264696, 21906767, 21906768, 55811957, 35695917, 265020, 264691, 264693, 27486262, 264628, 18108374, 35698423, 35695855, 264632, 264634, 264635, 264636, 264558, 18108384, 87168518, 22279000, 22279002, 264482, 284593, 264585, 264588, 264488, 29331822, 265007, 265010, 265018, 264769, 55811576, 56182323
2563	88095497 (5125, 5126)	Novel Protein sim. GBank gij4896447[emb]CAB43371.1] - (AL050270) hypothetical protein [Homo sapiens]		UNCLASSIFIED	
2564	80502783 (5127, 5128)	Novel Protein sim. GBank gij1352844[sp]P47179[JGP_YEAST - HYPOTHETICAL 118.4 KD PROTEIN IN BAT2-DALS INTERGENIC REGION PRECURSOR		sulfotransferase	
2565	85530906 (5129, 5130)			UNCLASSIFIED	66714117, 264609, 263978, 264632
2566	80224856 (5131, 5132)	Novel Protein sim. GBank gij628012[pr]A53933 - myosin I myr 4 - rat	Contains protein domain (PF00053) - Myosin head (motor domain)	struct	18108370, 35695855, 264558, 264558, 18108383
2567	86143590 (5133, 5134)			synthase	265020, 60170615
2568	81233099 (5135, 5136)	Novel Protein sim. GBank gij466009[sp]P34548[YNJ4_CAEEL - HYPOTHETICAL 33.8 KD PROTEIN R10E11.4 IN CHROMOSOME III			60424178, 18108394, 56181686, 56994075, 22278998, 264490, 264259, 29331822, 56182181, 29331824, 60424269, 29331825, 29331826, 29331828, 264509, 29331830, 265007, 265008, 265009, 33657402, 265010, 265011, 265018, 264448, 264683, 264288, 18108354, 264768, 21906766, 21906767, 35695917, 265021, 33657023, 18108362, 33657108, 33657182, 35695763, 60431528, 55810764, 18108379, 83373044, 18108385, 60432113, 284482

2569	95313764 (5137, 5138)	Novel Protein sim. GBank gi 2598560 gb AAB84166.1  - (AF029674) basic leucine zipper protein LZIP [Homo sapiens]	Contains protein domain (PF00170) - bZIP transcription factor	transcription factor	18108394, 56182575, 56181686, 22278995, 22278996, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 56182181, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 264905, 264906, 56182435, 265008, 265009, 264910, 60432228, 264592, 60433356, 60433438, 21808754, 87168559, 265017, 265018, 265019, 264682, 264448, 264288, 21906765, 21906766, 21906767, 21906768, 29148627, 21905769, 35695917, 265021, 265022, 52644150, 264690, 264691, 264692, 264693, 65274620, 263967, 35695763, 20281089, 263974, 18108374, 55810764, 35696423, 35695855, 264558, 18108381, 56182323, 18108382, 83373044, 18108385, 56526486, 87168516, 22279000, 22279002, 264666
2570	94138754 (5139, 5140)	Novel Protein sim. GBank gi 4758954 ref NP_004567.1 pPPP2 - protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), beta isoform	Contains protein domain (PF01240) - Protein phosphatase 2A regulatory subunit PR55	phosphatase	22278998, 28331822, 29331824, 66714117, 28331825, 60432289, 29331827, 35696052, 264907, 264510, 265007, 265009, 264758, 33109954, 265019, 264686, 264689, 265020, 265021, 264691, 264692, 264693, 18108374, 35695855, 264634, 56182323, 264639, 60170394, 83373044, 22276002, 264482
2571	87733750 (5141, 5142)	Novel Protein sim. GBank gi 732218 sp P34609 YO80_CAEEL - HYPOTHETICAL 128.6 KD PROTEIN ZK1098.10 IN CHROMOSOME III		stud	284508, 264905, 264907, 264828, 18108351, 284555, 264556, 284557, 284558, 284559
2572	87627560 (5143, 5144)	Novel Protein sim. GBank gi 488431 gmo CAB43260.1  - (AL050084) hypothetical protein [Homo sapiens]		UNCLASSIFIED	35696286, 29331827, 35696052, 264100, 264104, 264110, 264592, 21806754, 29148627, 29146629, 263972, 263974, 18108374, 263976, 35695855, 60170394, 264558, 18108385

2573	95313929 (5145, 5148)	Novel Protein sim. GBank gil399138 sp P02745 C1QA_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, A CHAIN PRECURSOR	Contains protein domain (PF00386) - C1q domain	complement	264488, 60424179, 65274572, 56182576, 56181686, 22278995, 56994075, 22278997, 60432049, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 264104, 264107, 264508, 264906, 29331830, 264909, 264510, 265008, 264512, 265008, 265009, 264910, 264591, 264592, 60432229, 264593, 60433356, 264594, 60433438, 264595, 55812038, 264759, 21906754, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 264761, 264762, 264763, 264764, 264369, 264288, 264685, 264766, 264686, 264687, 264688, 264769, 55181562, 264689, 21908785, 21908788, 21908789, 29148627, 21908788, 21908789, 265020, 265021, 265022, 60170615, 264689, 52844150, 264691, 264692, 33657023, 65274620, 18108365, 18108368, 27486265, 60431602, 264629, 60431528, 263976, 65274791, 35695855, 20281071, 60431850, 264637, 264638, 264558, 264639, 56182323, 60170394, 83373044, 18108384, 87188518, 80432113, 264482, 264564, 264565, 264568, 264567
2574	94746814 (5147, 5148)	Novel Protein sim. GBank R27216_1 [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	UNCLASSIFIED	22278995, 264259, 60432289, 29331827, 29331828, 33656970, 264908, 265008, 264910, 264591, 33657402, 265018, 265019, 264448, 264764, 264389, 264288, 18108357, 21908785, 21908788, 21908789, 55811857, 60170615, 264691, 33657023, 264693, 33657109, 33657182, 27486261, 27486264, 33657349, 264636, 264555, 63373044, 18108385, 264482
2575	87754408 (5149, 5150)	Novel Protein sim. GBank gil4929728 gb AAD34125.1 AF15188 - (AF15188) CGI-130 protein [Homo sapiens]		UNCLASSIFIED	264910, 264601, 264683, 264689, 264080
2576	95357881 (5151, 5152)	Novel Protein sim. GBank gil4680681 gb AAD27730.1 AF132955 - (AF132955) CGI-21 protein [Homo sapiens]	Contains protein domain (PF00442) - Ubiquitin carboxyl-terminal hydrolases family 2	ubiquitin	264259, 35686052, 264906, 60433438, 264881, 18108351, 264288, 52844150, 264828, 35696423
2577	86996621 (5153, 5154)	Novel Protein sim. GBank gil4337103 gb AAD18079 - (AF128756) NO26 [Homo sapiens]		UNCLASSIFIED	29331825, 265018, 265019, 264685
2578	87766941 (5155, 5156)			UNCLASSIFIED	264488, 264808, 264908, 264910, 264596, 264603, 264604, 264605, 264766, 21908789, 264828, 264630, 264634, 264639, 264563
2579	87282879 (5157, 5158)			UNCLASSIFIED	29331822, 29331824, 264767

2580	88166788 (5159, 5160)	Novel Protein sim. GBank gij2588628 (AC003080) - Similar to KIAA0299; 60% similarity to AB002297 (PID:g224539) [Homo sapiens]			265007, 265018, 264762
2581	87899048 (5181, 5182)	Novel Protein sim. GBank gij4406642 [gb AAD20049] - (AF131809) Unknown [Homo sapiens]		Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF).	56984075, 29331824, 29331826, 29331828, 264905, 60433356, 60433438, 264758, 87188599, 21906769, 265022, 35698855, 263881
2582	87786789 (5163, 5164)	Novel Protein sim. GBank gij2739367 (AC002505) - putative phosphatidylinositol-4-phosphate 5-kinase [Arabidopsis thaliana]		eph	264488, 264907, 264908, 264910, 264764, 264684, 264786, 264636, 264555, 264565
2583	81220950 (5165, 5166)	Novel Protein sim. GBank gij4378112 [emb CAA16521.1] - (AL021578) dJ453C12.2 (similar to transcription factor RBP-L) [Homo sapiens]		transcription factor	56181686, 264259, 264510, 264512, 264591, 264592, 264594, 264595, 264596, 264603, 264629, 55810764, 264630, 264637, 264565
2584	80430941 (5167, 5168)			UNCLASSIFIED	264908, 264910, 264768, 264693, 18108374, 55811576, 56182323
2585	80436126 (5169, 5170)	Novel Protein sim. GBank gij2736151 (AF021935) - mylonic dystrophy kinase-related Cdc42-binding kinase [Rattus norvegicus]		kinase	264768
2586	91226136 (5171, 5172)				22278998, 264259, 29331822, 29331824, 29331827, 29331828, 264905, 265007, 265009, 264591, 60433356, 33657402, 265018, 264782, 264288, 21806768, 21906767, 21906769, 265022, 264691, 83373044, 56526486, 22279002
2587	80430943 (5173, 5174)				264908, 265019, 264768, 264693, 55811576, 56182323
2588	80074385 (5175, 5176)			UNCLASSIFIED	264564
2589	85515607 (5177, 5178)	Novel Protein sim. GBank gij0021598 [emb CAA71415] - (Y10389) nuclear protein [Xenopus laevis]		UNCLASSIFIED	35688032, 284905, 284906, 284807, 284908, 264908, 265009, 265018, 264769, 35698423, 264636
2590	87054528 (5179, 5180)	Novel Protein sim. GBank gij2104689 (U92793) - alpha glucosidase II, alpha subunit [Mus musculus]		glucosylase	22278995, 29331830, 265008, 265010, 265017, 264639
2591	84192167 (5181, 5182)	Novel Protein sim. GBank gij5702202 [gb AAD47199.1] (AF12916 - (AF129166) long-chain acyl-CoA synthetase 5 [Homo sapiens])		eph	264259, 29331822, 264106, 264906, 56182435, 265007, 265008, 33109954, 264448, 55811957, 265020, 18108370, 55811576, 22278002

2592	95332648 (5183, 5184)	Novel Protein sim. GBank gi 3024998 sp Q60935 YAB1_MOUSE - HYPOTHETICAL HEART PROTEIN		transport	18108397, 56182575, 35696286, 56994075, 264259, 29331822, 29331824, 29331826, 60432289, 29331827, 29331828, 264906, 264909, 265007, 265008, 264810, 60432229, 264594, 60433356, 60433438, 55612038, 18108348, 21906754, 285011, 87166559, 265017, 265019, 264764, 264369, 264288, 264765, 265021, 60170615, 33657023, 33657109, 264629, 35696423, 35695855, 264557, 264638, 60170394, 56182323, 83373044, 56526486, 87168518, 264563, 264482, 264565
2593	87754416 (5185, 5186)	Novel Protein sim. GBank gi 4929729 gb AA034125.1 AF15188 - (AF15188) CGI-130 protein [Homo sapiens]	tm7		22278999, 29331825, 264758, 21906754, 52846317, 265010, 18108351, 264288, 264369, 21906768, 264693, 18108370, 264637, 264638, 264482
2594	95305758 (5187, 5188)	Novel Protein sim. GBank gi 4929587 gb AA034054.1 AF15181 - (AF15181) CGI-59 protein [Homo sapiens]	UNCLASSIFIED		264488, 18108398, 56182575, 35696286, 22278997, 264093, 264259, 29331822, 29331825, 68714117, 29331826, 264905, 264909, 52644045, 56182435, 264510, 264512, 265007, 264757, 21906754, 87168474, 285017, 264760, 264448, 264764, 264288, 264766, 264889, 21906768, 33657109, 263975, 263977, 264634, 264556, 60170394, 56182323, 56526486, 264482, 264563, 264564, 264566, 264567
2595	79561676 (5189, 5190)	Novel Protein sim. GBank gi 430968 gb AA015478  - (AC006930) R33423_1 [Homo sapiens]	UNCLASSIFIED		264692
2596	87538637 (5191, 5192)	Novel Protein sim. GBank gi 430968 gb AA015478  - (AC006930) R33423_1 [Homo sapiens]	UNCLASSIFIED		22278999, 264259, 265018, 264448, 265021, 60431528
2597	94784089 (5193, 5194)	Novel Protein sim. GBank gi 100135 dbj BAA10838  - (D64006) hypothetical protein [Synechocystis sp.]	UNCLASSIFIED		264905, 264509, 264908, 264762, 264766, 35695917, 35695855, 264635, 264636, 83373044, 264486
2598	88094948 (5195, 5196)	Novel Protein sim. GBank gi 100135 dbj BAA10838  - (D64006) hypothetical protein [Synechocystis sp.]	UNCLASSIFIED		22278998, 264259, 29331824, 87168474, 264683, 21906766, 35695917, 264691, 33657023, 33657109, 18108370, 18108374, 264564, 264565
2599	87842888 (5197, 5198)	Novel Protein sim. GBank gi 3941737 (AF109719) - BAT2 [Mus musculus]	MHC		264768, 264769, 21906768, 33657182, 35695763, 18108370, 18108374, 264635, 264636, 56526486, 22279000, 264566
2600	87787848 (5199, 5200)	Novel Protein sim. GBank gi 426352 gb AA015347  - (AC004044) putative WD-repeat protein [Arabidopsis italiana]	Contains protein domain (PF00400) - WD domain, G-beta repeat		35696286, 264093, 264288, 21906769, 35696423, 35695855

2601	81243070 (5201, 5202)	Novel Protein sim. GBank gij728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		kinase	56182575, 22278999, 264256, 29331822, 29331824, 60432289, 29331826, 29331827, 264906, 265007, 285008, 264591, 60433356, 33657402, 60433438, 21906754, 265011, 265018, 265019, 18108351, 264448, 264369, 21906769, 265020, 60170615, 264693, 33657109, 18108370, 18108376, 56182323, 18108381, 18108395, 22279002, 264563, 60433438, 21906754, 87168559, 264601, 264369, 264288, 21906767
2602	88180022 (5203, 5204)	Novel Protein sim. GBank gij406632 p AAD200471 - (AF131801) Unknown [Homo sapiens]		UNCLASSIFIED	264488, 65274572, 22278995, 22278996, 56994075, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 60432289, 29331826, 29331827, 29331828, 35896032, 56182435, 264113, 265008, 285008, 60433356, 264757, 60433438, 264759, 33657084, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 264448, 264683, 18108354, 264288, 264767, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265020, 265021, 265022, 60170615, 264691, 33657023, 264693, 33657109, 27486262, 18108374, 35896423, 85274791, 35895855, 264555, 264636, 264637, 56182323, 83373044, 56526486, 87168518, 60432113, 22279000
2604	94876601 (5207, 5208)	Novel Protein sim. GBank gij5454030 ref NP_006468.1 pRRP2 - RAS-related on chromosome 22		oncogene	264259, 35896052, 264508, 264906, 264907, 264908, 264909, 264510, 264512, 265008, 264910, 33657402, 264604, 264605, 264762, 264763, 264682, 264764, 264683, 264788, 264789, 264689, 33657023, 264693, 18108365, 264628, 35896423, 264631, 264632, 264634, 264635, 264637, 18108381, 264639, 83373044, 264565
2605	94316756 (5209, 5210)	Novel Protein sim. GBank gij3628745 db BAA333561 - (AB013721) mltaugmin 23 [Oryctolagus cuniculus]		UNCLASSIFIED	22278998, 264490, 60432049, 264259, 60432289, 264909, 265008, 60433356, 60433438, 264758, 21906754, 265010, 265011, 265018, 264681, 18108351, 264288, 264766, 264685, 21906765, 21906766, 21906768, 21906769, 264691, 264692, 264693, 85274791, 264634, 264555, 264636, 22278996, 264510, 264512, 265009, 264766, 22279002, 264566
2606	87746406 (5211, 5212)				

2607	87627742 (5213, 5214)	Novel Protein sim. GBank gi 4826626 gb AAD30202.1  - (AF135022) mediator [Homo sapiens]			29331822, 29331825, 29331826, 29331827, 29331828, 264908, 264907, 264908, 66712502, 264828, 56182435, 55812038, 265010, 265017, 265018, 265019, 264768, 264689, 21906765, 55811957, 265020, 265022, 264892, 33657023, 264693, 33657109, 18108370, 264639, 56182323, 264509, 264807, 264908, 264592, 264758, 264631
2608	81734768 (5215, 5216)	Novel Protein sim. GBank gi 2226005 (U49973) - ORF2: function unknown [Homo sapiens]			
2609	94843791 (5217, 5218)	Novel Protein sim. GBank gi 3024889 sp P56524 Y288_HUMAN - HYPOTHETICAL PROTEIN KIA0288 (HA6116)	Contains protein domain (PF00850) - Histone deacetylase family	histone	264488, 65274572, 35696286, 22278997, 22278999, 60432049, 264259, 56182181, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 264905, 264907, 60433356, 60433438, 55812038, 265011, 87168559, 265017, 265018, 264448, 264765, 264288, 264766, 264689, 21906765, 21906787, 21906789, 265020, 265021, 264691, 264692, 33657109, 27485261, 18108370, 65274791, 264636, 264596, 56182323, 18108385, 56526486
2610	88177654 (5219, 5220)	Novel Protein sim. GBank gi 4336855 gb AAD17989  - (AF106473) leucine-rich-domain inter-acting protein 1; LeR Inter-acting protein 1; LEAP1 [Mus musculus]		transcript factor	18108384, 22278994, 56994075, 60432049, 264259, 29331822, 29331825, 60432289, 29331827, 264107, 264109, 264905, 56182435, 264112, 265006, 265007, 265008, 265009, 60433356, 60433438, 265011, 87168559, 265017, 264448, 264682, 264784, 264288, 265021, 33657023, 263967, 33657182, 27488281, 18108374, 263978, 55811576, 264636, 87168518, 60432113, 22278999, 265017, 264684, 21906788, 22279000
2611	87428890 (5221, 5222)	Novel Protein sim. GBank gi 387676 emb CAA92994  - (Z68760) predicted using GeneFinder: Similarity to Mouse FKBP-type peptidyl-prolyl cis-trans isomerases	Contains protein domain (PF00254) - FKBP-type peptidyl-prolyl cis-trans isomerases	isomerase	
2612	87771198 (5223, 5224)	Novel Protein sim. GBank gi 5879138 gb AAD46874.1 AF16093 - (AF160934) BcDNA LD14189 [Drosophila melanogaster]		transport	265009, 264910, 264759, 265017, 21906767, 18108385, 18108388, 60432113
2613	79481496 (5225, 5226)			UNCLASSIFIED	264685
2614	87643948 (5227, 5228)	Novel Protein sim. GBank gi 5533081 gb AAD45009.1 AF16118 - (AF161181) P55T protein [Mus musculus]	Contains protein domain (PF00625) - Guanylate kinase		22278998, 22278999, 29331825, 264508, 264906, 21906754, 264602, 264766, 264769, 52644229, 21906765, 33657109, 27488264, 18108370, 283972, 264555, 60432113
2615	87381996 (5229, 5230)			UNCLASSIFIED	264788, 18108394, 264692, 264693, 264508, 264509, 264907, 264628, 264908, 264909, 18108377, 264511, 264512, 264910, 264635, 264595, 265010, 264404, 264563, 264764, 264685, 264768

2616	87428895 (5231, 5232)	Novel Protein sim. GBank gl 387676 emb CAA92894  - (Z68760) predicted using GeneFinder. Similarity to Mouse FK506-binding protein (SW:FKB_MOUSE) [Caenorhabditis elegans]	Contains protein domain (PF00254) - FKBP-type peptidyl-prolyl isomerases	isomerase	22278995, 22278997, 22278998, 60432049, 60432289, 264828, 60433356, 264594, 60433438, 33109954, 87168474, 265011, 265017, 265019, 264288, 264766, 21908765, 21908767, 265020, 265021, 18108378, 18108377, 18108387, 87168518, 264482, 264567
2617	86976888 (5233, 5234)	Novel Protein sim. GBank gl 728831 sp P39188 ALU1_HUMAN - IIII ALU SUBFAMILY J WARNING ENTRY!!!!		kinase	265010, 265019, 264369, 264693, 55811576, 22279002
2618	91231662 (5235, 5236)	Novel Protein sim. GBank gl 3319282 (AF049103) - Huntingtin interacting protein [Homo sapiens]	Contains protein domain (PF00397) - WW domain	UNCLASSIFIED	284489, 22278996, 284490, 284259, 29331822, 264102, 284506, 264908, 264907, 68712502, 29331830, 265008, 264910, 265009, 60433356, 60433438, 264758, 21908754, 265011, 87168559, 265017, 265018, 264369, 264288, 264766, 264768, 264688, 21908765, 21908766, 21908767, 35695917, 265020, 265022, 33657023, 264692, 33657109, 264628, 18108374, 35695855, 18108381, 83373044, 18108385, 18108388, 56526486, 264563
2619	87684000 (5237, 5238)	Novel Protein sim. GBank gl 2431772 (U56411) - putative type III alcohol dehydrogenase [Drosophila melanogaster]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenases	dehydrogenase	264259, 60432289, 60433438, 21908754, 264369, 60432113, 264566
2620	95314841 (5239, 5240)	Novel Protein sim. GBank gl 4322567 gb AAD16097  - (AF090438) dactshund variant 1 [Mus musculus]	Contains protein domain (PF00628) - PHD-finger	UNCLASSIFIED	52644507, 52645158, 52646842, 65274572, 22278995, 5694075, 35695286, 22278999, 60432049, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331828, 35696052, 264907, 66712502, 265008, 60433356, 33657402, 52646317, 21908754, 87168474, 265010, 265017, 265018, 264448, 264369, 264288, 264687, 264768, 52644229, 264688, 264689, 21908765, 21908768, 35695917, 52644150, 264692, 33657109, 35695763, 35696423, 264556, 52644332, 18108382, 83373044, 18108385, 18108387, 65274727, 87168518, 60432113, 22279002
2621	80253495 (5241, 5242)				264594, 264636
2622	81780390 (5243, 5244)	Novel Protein sim. GBank gl 4537341 ref NP_001174.1 pATP6 - ATPase, H <sup>+</sup> transporting, lysosomal subunit 1; vacuolar proton pump; H <sup>+</sup> ATPase subunit			284488, 284908, 284907, 284908, 284512, 285007, 264758, 35695917, 264634, 264636, 264583, 264482
2623	91639306 (5245, 5246)	Novel Protein sim. GBank gl 3860355 emb CAB05289  - (Z62285) predicted using GeneFinder [Caenorhabditis elegans]		UNCLASSIFIED	29331824, 35696032, 265007, 265010, 264288, 29148829



2624	91639308 (5247, 5248)	Novel Protein sim. GBank gi 3880355 emb CA05289  - (Z52285) predicted using GeneFinder (Caenorhabditis elegans)		UNCLASSIFIED	56181686, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 56182181, 29331825, 60432289, 29331828, 35698052, 29146499, 68712502, 52644045, 265007, 265008, 60433356, 33109954, 21908754, 265010, 265011, 265018, 264448, 264288, 21908765, 21906766, 21906767, 29148629, 35698917, 265021, 265022, 27486285, 18108370, 60431528, 55811576, 35698555, 58182323, 18108385, 87168518, 22279002, 18108391
2625	86452068 (5249, 5250)	Novel Protein sim. GBank gi 2887429 dbj BA024857  - (AB007887) KIAA0427 (Homo sapiens)		UNCLASSIFIED	264091, 264511, 263981
2626	16533797 (5251, 5252)	Novel Protein sim. GBank gi 487416 (L20302) - actin filament protein (Gallus gallus)		struct	265008
2627	87638823 (5253, 5254)	Novel Protein sim. GBank gi 88462 pir J27307 - proline-rich phosphoprotein (gene PRH1, Db allele) - human		UNCLASSIFIED	22278996, 265007, 265009, 264448, 21906767, 285021, 264558, 87168518
2628	94848254 (5255, 5256)	Novel Protein sim. GBank gi 323552 emb CAA18609  - (AL022578) dJ393P12.2 (hypothetical Proline-rich protein KIAA0269 LIKE) (Homo sapiens)		UNCLASSIFIED	22278997, 22278999, 264259, 60432049, 29331822, 29331824, 29331825, 29331827, 35698052, 29331828, 264907, 264909, 265008, 264591, 60433356, 60433438, 265010, 265017, 265018, 264369, 264288, 18108357, 21906765, 21906768, 265022, 65274791, 264638, 18108387, 87168518, 22279002
2629	87376490 (5257, 5258)	Novel Protein sim. GBank gi 4829595 gb AAD34058.1 AF15182 - (AF15182) CGI-63 protein (Homo sapiens)		synthase	29331825, 29331826, 264102, 265008, 264766, 35695917, 264691, 33557023, 263972, 18108374, 22279000
2630	79188364 (5259, 5260)	Novel Protein sim. GBank gi 321605 pir JQ1161 - Gag protein - Virus virus (strain EV1)		UNCLASSIFIED	264636, 18108385
2631	94845909 (5261, 5262)	Novel Protein sim. GBank gi 321605 pir JQ1161 - Gag protein - Virus virus (strain EV1)	Contains protein domain (PF00098) - dna_m_bind Zinc finger, CCHC class	UNCLASSIFIED	52644507, 52645156, 52646365, 52646842, 22278994, 22278995, 35698286, 56984075, 22278997, 22278999, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 35698052, 33656970, 264905, 264509, 264907, 264908, 264511, 264512, 265007, 265008, 264910, 52646317, 33657084, 52644296, 265010, 87168559, 265017, 265018, 265019, 264760, 264762, 264448, 264288, 264369, 264768, 264768, 52644229, 21906764, 21906785, 21908786, 21906767, 21906789, 35695917, 265020, 52644150, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 27486265, 33657348, 35698763, 35698423, 65274791, 35698555, 264634, 264637, 52644332, 58182323, 60432113, 264556, 264488
2632	36730414 (5263, 5264)				264695

2633	95011617 (5265, 5266)	Novel Protein sim. GBank gi11139548 dbj BAA10888  - (D64009) seizure-related gene product 8 type 2 precursor [Mus musculus]	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	22278995, 22278997, 22278998, 264259, 29331822, 29331824, 29331825, 29331827, 264508, 264509, 265006, 265007, 265008, 265009, 55812038, 33637084, 55811388, 265010, 265011, 87168559, 265018, 265019, 264683, 264288, 264686, 29148829, 33637023, 284693, 33657182, 35695763, 55811576, 264639, 58182323, 83373044, 18108385, 56528486, 87168518, 22279000, 22279002, 284595
2634	87330921 (5287, 5288)	Novel Protein sim. GBank gi15441611 emb CAB46854.1  - (AJ388555) hypothetical protein [Canis familiaris]	UNCLASSIFIED	22278997, 264259, 29331824, 56714117, 29331827, 29331828, 264907, 33657084, 265017, 265018, 264448, 264288, 21908768, 21908767, 21908768, 29148829, 18108376, 55811576, 35695855, 87168518, 22279000
2635	86823144 (5269, 5270)	Novel Protein sim. GBank gi4680663 gb AAD27721.1 AF132946  CGI-12 protein [Homo sapiens]		264569, 29331822, 29331828, 265006, 60170831, 264681, 264765, 264685, 29148827, 21908768, 29148784, 265022, 60170615, 264635, 18108385, 56528486, 22279002, 264567
2636	87280534 (5271, 5272)	Novel Protein sim. GBank gi3879146 emb CAB07546  - (Z93386) Similarity to Yeast hypothetical 52.9 KD protein (SWP43616); cDNA EST EMBL:M89432 comes from this gene; cDNA EST EMBL:D71008 comes from this gene; cDNA EST EMBL:D73578 comes from this gene; cDNA EST EMBL:D68025 comes ...	Contains protein domain (PF01546) - Peptidase family M20/M25/M40	264488, 264489, 52644507, 264487, 52646365, 52646842, 22278994, 22278998, 22278999, 20281171, 264289, 29331822, 52645080, 86714117, 29331825, 29331826, 29331827, 35696052, 29331828, 29146488, 264508, 264905, 264906, 264907, 264908, 264909, 264511, 264512, 264910, 264591, 264592, 60432228, 264583, 264594, 33657402, 60433356, 264757, 60433438, 264596, 264758, 52646317, 21906754, 52644296, 265010, 264600, 264602, 264603, 264605, 264761, 264762, 264681, 264448, 264764, 264765, 264288, 264768, 264686, 264768, 264687, 264769, 21906765, 21908768, 21906767, 21906768, 21906769, 35695917, 265021, 60170615, 264691, 33657023, 264692, 264693, 65274620, 27486284, 18108370, 284628, 264629, 18108374, 35696423, 35696855, 284632, 264634, 264635, 284637, 264638, 52644332, 264639, 264558, 83373044, 60432113, 264564, 264565, 264566, 264488, 264567
2637	95011298 (5273, 5274)	Novel Protein sim. GBank gi4758208 ref NP_004081.1 pDUSP - dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	Contains protein domain (PF00782) - Dual specificity phosphatase, catalytic domain	

2638	94326733 (5275, 5276)	Novel Protein sim. GBank gi 4929898 gb AAD34105.1 AF15186 - (AF15186) CGI-110 protein [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	60424179, 52644507, 52646842, 18108398, 56182575, 22278995, 22278998, 35698288, 22278997, 22278999, 264259, 60432049, 29331822, 60424289, 29331826, 35696052, 29146498, 264905, 52644045, 56182435, 60433356, 33657402, 55812038, 55811386, 265019, 264288, 284769, 52844229, 56181562, 29148827, 28148629, 55811857, 29148784, 35695917, 265021, 52844150, 33657023, 65274820, 33657109, 35695763, 18108374, 55810764, 35696423, 55811576, 35695855, 60431850, 56182323, 60432113, 2644404
2639	95381348 (5277, 5278)	Novel Protein sim. GBank gi 2180007 dbj BAA20355  - (AB004109) phosphatidylserine synthase II [Cricetus griseus]		synthase	264488, 29331825, 35696052, 264508, 264509, 284909, 264512, 33657402, 60433438, 264758, 85695342, 264600, 265020, 265021, 33657109, 284628, 35696423, 264555, 264639, 264563, 264564, 264565, 264566, 264486
2640	87781330 (5279, 5280)	Novel Protein sim. GBank gi 3158516 (AF067617) - contains similarity to chromo (chromatin organization modifier) domains (Pfam: chromo.hmm, score: 17.76 and 27.94) and to helicases conserved C-terminal domain (Pfam: helicase_C.hmm, score: 67.00) [Caenorhabditis elegans]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	29331822, 29331826, 264906, 33109954, 265017, 265019, 21906768, 35695763, 264636, 264637, 18108387
2641	11669834 (5281, 5282)	Novel Protein sim. GBank gi 2564955 (AF030001) - unknown [Mus musculus]		UNCLASSIFIED	264928
2642	87412575 (5283, 5284)			UNCLASSIFIED	264259, 29331822, 33657402, 265019, 264369, 264691, 264634, 56526486, 22279002
2643	87643981 (5285, 5286)	Novel Protein sim. GBank gi 4490304 emb CAC38795.1  - (AL035678) putative protein [Arabidopsis thaliana]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	22278997, 264259, 29146499, 56182435, 264910, 265010, 18108351, 264682, 264683, 264369, 264684, 264685, 264686, 29148627, 264690, 33657109, 18108370, 263973, 18108374, 264634, 264557, 264558, 18108385, 264482
2644	88177671 (5287, 5288)	Novel Protein sim. GBank gi 3769797 gb AAC67502.1  - (AF059569) actin binding protein MAYVEN [Homo sapiens]	Contains protein domain (PF00851) - BTB/POZ domain	nuc_l_rept	264107, 264687
2645	17277228 (5289, 5290)			UNCLASSIFIED	265007
2646	94148542 (5291, 5292)	Novel Protein sim. GBank gi 1708722 sp P49749 EVX2_MOUSE - HOMEBOX EVEN SKIPPED HOMOLOG PROTEIN 2 (EVX-2)		UNCLASSIFIED	264909, 264687, 264632, 63373044

2647	91212878 (5283, 5284)			UNCLASSIFIED	56182575, 22278986, 35696286, 22278988, 264259, 29331822, 56182181, 29331825, 60424268, 60432289, 35698052, 68712502, 264908, 265007, 55812038, 33109954, 21906754, 33857084, 265019, 264448, 284288, 56181562, 21906765, 21906768, 21908768, 21908769, 35695917, 265020, 265021, 52844150, 264693, 33657109, 33657349, 60431528, 18108374, 55810764, 35696423, 56182323, 60432113, 22278002, 264584
2648	87600587 (5295, 5296)				29148468, 56182435, 33109954, 265011, 264682, 55811957, 35695917, 264690, 263978, 18108377, 35696423, 60432113
2649	94128783 (5297, 5298)	Novel Protein sim. GBank gij3041852 (AC004538) - unknown function, similar to Y09105 (PID.g1666171) [Homo sapiens]	UNCLASSIFIED		56182575, 35696288, 22278988, 29331824, 29331826, 60432289, 68712502, 56182435, 60170831, 60432228, 33857402, 33109954, 21906754, 265017, 264688, 284688, 21906765, 21906768, 60170815, 264693, 263987, 18108370, 263976, 60170394, 60432113, 22278002, 264553, 264685
2650	87297533 (5299, 5300)	Novel Protein sim. GBank gij5360271dbj[BAA81908.1] - (AB029335) HRPET-3 [Halocynthia roretzi]			
2651	88088745 (5301, 5302)	Novel Protein sim. GBank gij4240225dbj[BAA74891.1] - (AB020675) KIAA0868 protein [Homo sapiens]	Contains protein domain (PF00054) - Laminin G domain	- synthase	29331824, 29331826, 29331827, 265007, 55812038, 21906754, 18108366, 18108384, 22278002, 284587
2652	10343125 (5303, 5304)			UNCLASSIFIED	264692
2653	87798735 (5305, 5306)	Novel Protein sim. GBank gij44939961emb[CAB11123.2] - (Z98551) predicted using hexExon; MAL3P6.28 (PFC0845c). Hypothetical protein, len: 167 aa. Similarity to model organism hypothetical proteins (C.elegans, D.melanogaster, S.cerevisiae & S.pombe). C.elegans protein ZK287.5 (TR....		UNCLASSIFIED	265018, 18108370, 18108387, 264556
2654	65103240 (5307, 5308)				60424178, 65274572, 56182575, 264259, 56182181, 284908, 56182435, 55811957, 35695917, 265021, 283976, 55810764, 65274781, 56182323, 83373044, 65274727
2655	91228018 (5309, 5310)	Novel Protein sim. GBank gij38752721emb[CAB02861] - (Z81051) predicted using GeneFinder; similar to Zinc finger, C3HC4 type (RING finger); cDNA EST yk443h5.3 comes from this gene; cDNA EST yk443h5.5 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	- transcription factor	56182575, 56181686, 284902, 264259, 56182181, 60432289, 284907, 33657402, 55812038, 21906754, 87168559, 265017, 264448, 264369, 264286, 21906765, 21906768, 21906767, 21906768, 33657109, 18108370, 264628, 55811578, 264556, 264639, 83373044, 56528488, 264404, 60432113
2656	94562501 (5311, 5312)	Novel Protein sim. GBank gij3043718dbj[BAA25523] - (AB011169) KIAA0597 protein [Homo sapiens]			264683

2657	52561728 (5313, 5314)	Novel Protein sim. GBank gi 5689509 dbj BAA63038.1  - (AB029009) KIAA1086 protein [Homo sapiens]		dna_rna_bind	264693
2658	88062454 (5315, 5316)	Novel Protein sim. GBank gi 3688089 (AC005757) - R32611_1 [Homo sapiens]	Contains protein domain (PF00360) - Leucine Rich Repeat	nucleaseinhib	35696286, 264259, 29331822, 29331824, 29331826, 29331828, 265018, 264683, 21906768, 35695917, 264693, 35695855, 264637, 87108518, 264488, 264587, 264909, 264910, 265018, 264369, 264769, 21906769, 264693, 263972, 18108388
2659	87600755 (5317, 5318)	Novel Protein sim. GBank gi 5420387 emb CA6679.1  - (AJ243459) proteophosphoglycan [Leishmania major]	Contains protein domain (PF01426) - BAH domain	UNCLASSIFIED	264909, 264910, 265018, 264369, 264769, 21906769, 264693, 263972, 18108388
2660	81718472 (5319, 5320)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7 - SQ WARNING ENTRY IIII	Contains protein domain (PF00036) - EF hand	kinase	264488, 65274572, 35696286, 22278908, 22278998, 264259, 29331822, 29331824, 29331826, 29331828, 35696052, 264908, 60432289, 29331828, 35696052, 264908, 56182435, 265008, 265009, 60433356, 264594, 265010, 265018, 59811150, 18108351, 264682, 264684, 264369, 264288, 264687, 21906765, 26148784, 35695917, 60170615, 52844150, 33657023, 33657109, 35696423, 35695855, 264556, 60170394, 18108385, 22279000, 22279002
2661	95342817 (5321, 5322)	Novel Protein sim. GBank gi 4758048 ref NP_004739.1 pCPR8 - cell cycle progression 8 protein		glycoprotein	60432049, 264259, 29331824, 29331825, 29331826, 29331827, 29331828, 264908, 264909, 264593, 33109954, 265010, 265017, 265018, 265019, 264760, 264448, 264369, 264288, 21906765, 21906768, 265022, 264691, 33657023, 27488282, 60431528, 18108374, 35695855, 18108386, 264482, 264555, 264556, 264558, 264486
2662	80228739 (5323, 5324)	Novel Protein sim. GBank gi 3874714 emb CAA91263  - (Z66494) similar to choline dehydrogenase. cDNA EST yk346d5.5 comes from this gene; cDNA EST yk346d5.3 comes from this gene [Caenorhabditis elegans]		dehydrogenase	264906, 264909, 264757, 264758, 264767, 264691, 33657023, 264638
2663	87760823 (5325, 5326)	Novel Protein sim. GBank gi 3874714 emb CAA91263  - (Z66494) similar to choline dehydrogenase. cDNA EST yk346d5.5 comes from this gene; cDNA EST yk346d5.3 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00008) - EGF-like domain	oncogene	35696286, 264508, 264595, 264288, 264685, 264686
2664	85518329 (5327, 5328)	Novel Protein sim. GBank gi 1389670 (U56977) - Notch homolog Scalloped wings [Lucilia cuprina]		UNCLASSIFIED	35696286, 22278999, 29331822, 35696052, 264906, 264907, 264909, 264510, 264511, 264512, 264593, 60433438, 265018, 264681, 21906765, 21906766, 21906767, 21906768, 265020, 265022, 35696423, 35695855, 22279002, 264482, 264486
2665	87770662 (5329, 5330)	Novel Protein sim. GBank gi 4884408 emb CA843311.1  - (AL050190) hypothetical protein [Homo sapiens]		UNCLASSIFIED	29331825, 285007, 264910, 60432229, 265018, 264288, 21906767, 264558, 22279002
2666	87826472 (5331, 5332)	Novel Protein sim. GBank gi 5108956 gb AA039906.1 AF113615 - (AF113615) FH1/FH2 domain-containing protein FHO5 [Homo sapiens]			264907, 29331830, 264681, 264683, 264288, 35695855, 264632, 264556, 264557, 264558, 264559, 264583, 264585, 264587
2667	87422720 (5333, 5334)	Novel Protein sim. GBank gi 2500570 sp Q17533 RNPH_CAEEL - RIBONUCLEASE PH-LIKE PROTEIN B0584.1	Contains protein domain (PF01138) - 3' exoribonuclease family	nuclease	

2668	91216716 (5335, 5336)	Novel Protein sim. GBank gi 5454186 ref NP_006327.1 pZYG  - ZYG homolog	UNCLASSIFIED	56181886, 35698286, 22278988, 22278989, 56182181, 29331824, 60424269, 29331825, 35696052, 29331828, 66712502, 56182435, 60433358, 264758, 21906754, 55811386, 265011, 87186559, 265017, 265019, 55811150, 264448, 264369, 264288, 21906765, 21906766, 21906768, 55811957, 35695917, 265020, 265021, 33657023, 264682, 33657109, 35695763, 60431828, 18108374, 35696423, 55811576, 264634, 60431850, 83373044, 18108385, 87168518, 22279000, 264563, 264564
2669	95415721 (5337, 5338)	Novel Protein sim. GBank gi 2147012 pri JC4899 - proline rich protein - rat	UNCLASSIFIED	254489, 264689, 21906767, 65274572, 56182575, 21906768, 29148627, 21906769, 29148629, 35696286, 35695917, 22278988, 22278989, 285021, 285022, 60170615, 52644150, 60432048, 264258, 264691, 33657023, 264692, 29331822, 29331824, 29331825, 60432289, 33657109, 29331826, 29331827, 35696052, 29331828, 29146488, 29146499, 264905, 264908, 52644045, 264909, 56182435, 35696423, 65274781, 35695865, 265006, 264810, 264835, 60432229, 264592, 264638, 56182323, 60433356, 60170394, 60433438, 264559, 264595, 55812038, 33109954, 87168559, 60432113, 265019, 264448, 264369, 264684, 264288
2670	87610234 (5339, 5340)	Novel Protein sim. GBank gi 1723523 sp Q10362 YD8B_SCHPO - HYPOTHETICAL 94.9 KD PROTEIN C22E12.11C IN CHROMOSOME I	Contains protein domain (PF00628) - PHD-finger	18108370, 263974
2671	91214936 (5341, 5342)	Novel Protein sim. GBank gi 476827 gb AAD28444.1 AF06425 - (AF06425) very long-chain acyl-CoA synthetase homolog 2; VLCS-H2 [Homo sapiens]	transport	52646842, 56994075, 264259, 29331822, 29331824, 29331825, 29331827, 33658970, 284508, 265006, 33109954, 21906754, 264682, 264288, 265021, 33657023, 33657109, 33657182, 27486261, 27486262, 27486285, 18108376, 18108385
2672	87399123 (5343, 5344)	Novel Protein sim. GBank gi 4966348 gb AAD34677.1 AC00834 - (AC00834) Contains two PF01344 Kelch motif domains. [Arabidopsis thaliana]	UNCLASSIFIED	264767

2673	87430749 (5345, 5348)	Novel Protein sim. GBank gij5457337[emb]CAB41505.2] - (AJ236876) poly(ADP-ribose) polymerase-2 [Homo sapiens]	Contains protein domain (PF00644) - Poly(ADP-ribose) polymerase catalytic region.	polymerase	22278995, 22278996, 22278997, 22278998, 29331822, 29331824, 29331828, 35696052, 60433438, 87168474, 87168559, 265017, 265018, 265019, 264448, 21908788, 21908789, 265020, 265021, 33657109, 27488282, 35695763, 60431850, 60170394, 87168518, 284563
2674	94847721 (5347, 5348)	Novel Protein sim. GBank gij4758824[ref]NP_004280.1[pNRF3 - nuclear factor (erythroid-derived 2)-like 3]	Contains protein domain (PF00170) - bZIP transcription factor	transcription factor	264488, 22278996, 35696286, 264091, 264259, 29331824, 29331826, 35696052, 264511, 55812038, 85658542, 264766, 21908785, 35695917, 264628, 35696423, 18108383, 87168518
2675	79563835 (5348, 5350)			UNCLASSIFIED	264691
2676	79628383 (5351, 5352)			UNCLASSIFIED	264906, 265008
2677	94328600 (5353, 5354)	Novel Protein sim. GBank gij1079042[pi][S52154 - acetyl-CoA synthetase - fruit fly (Drosophila melanogaster)]	Contains protein domain (PF00501) - AMP-binding enzyme	synthase	18108394, 65274572, 56182575, 22278994, 22278995, 56994075, 22278996, 35696286, 22278997, 22278999, 264490, 264259, 52645080, 29331824, 29331825, 60432289, 29331827, 29331828, 35698052, 28148489, 29331830, 264908, 52644045, 265006, 265007, 265008, 265009, 60432229, 60433356, 60433438, 55812038, 265010, 265011, 87168559, 265017, 265019, 18108351, 264682, 264448, 264683, 264288, 21908765, 21908766, 21908787, 21908788, 21908789, 55811957, 265020, 265021, 60170615, 52644150, 264691, 33657023, 265967, 33657109, 27488284, 27488285, 33657348, 35695763, 18108370, 18108374, 18108377, 55811576, 35696423, 35695955, 83373044, 18108387, 22279000, 22279002, 264584

2878	95001894 (5355, 5356)	Novel Protein sim. GBank gijl85760 prijA0465 - alanine transaminase (EC 2.6.1.2), cytosolic - human		UNCLASSIFIED	264488, 263994, 264489, 18108394, 52646842, 35696286, 22278999, 264259, 29331823, 35898052, 284508, 284909, 284509, 284908, 284907, 284908, 284909, 284510, 284511, 285008, 264512, 265007, 265008, 264910, 265009, 264593, 264594, 264596, 264758, 33109954, 52646317, 21906754, 265010, 265011, 87168559, 264600, 264601, 264602, 264603, 265017, 265019, 264605, 264760, 264762, 18108351, 264763, 264682, 264683, 264764, 264288, 264766, 284887, 284768, 284769, 284688, 21808765, 21908787, 35695917, 285020, 265021, 52844150, 264691, 33657023, 264693, 18108364, 18108365, 33657109, 33657349, 284628, 18108374, 35696423, 35695955, 284630, 284631, 284632, 264634, 264635, 264555, 264636, 264637, 264638, 264639, 18108385, 264553, 264584, 264565, 264586, 284488, 284587
2879	95361544 (5357, 5358)	Novel Protein sim. GBank gijl709233 spjP07514 NCSR_BOVIN - NADH- CYTOCHROME B5 REDUCTASE	Contains protein domain (PF00970) - FAD/NAD-binding Cytochrome reductase	UNCLASSIFIED	264488, 22278996, 35696286, 264259, 29331826, 29331827, 29331828, 284909, 56182435, 264113, 264511, 265008, 60433438, 284758, 85658542, 87168474, 265011, 285017, 265019, 284288, 21908766, 21908787, 21908788, 21908789, 55811857, 265021, 264690, 33657023, 55810764, 35696423, 55811576, 284631, 18108381, 60170394, 83373044, 87168518, 264566
2680	87800356 (5359, 5360)	Novel Protein sim. GBank gij4589504 dbjBAA76824.1  - (AB023197) KIAA0980 protein [Homo sapiens]	Contains protein domain (PF00036) - EF hand	UNCLASSIFIED	264259, 264102, 264905, 264908, 265007, 265008, 33109954, 285011, 18108351, 264788, 33657023, 20281148, 283872, 264630, 284635, 284638
2681	80933844 (5361, 5362)	Novel Protein sim. GBank gijl728637 spjP09194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII	kinase	UNCLASSIFIED	284489, 22278995, 284509, 284905, 284906, 284907, 284908, 284909, 264510, 265008, 284512, 284910, 284594, 60433438, 264758, 284603, 284604, 265019, 284605, 264760, 284764, 284687, 264768, 284769, 21908769, 35695917, 285020, 33657023, 264631, 284635, 284637, 284638, 284639, 284488
2682	84138934 (5363, 5364)	Novel Protein sim. GBank gij423468 prijJQ1974 - HTF9-C protein - mouse	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	35696286, 284908, 55811388, 285017, 55811150, 55811857, 35695917, 60431528, 55810764, 55811576, 35696423, 65274791, 56326486
2683	87774405 (5365, 5366)	Novel Protein sim. GBank gij5114351 g0jAAD40286.1  - (AF156271) RING finger protein terf [Homo sapiens]	Contains protein domain (PF00622) - SPRY domain	interleukinrecept	284909, 284769, 264635, 264636



2684	85787151 (5387, 5388)	Novel Protein sim. GBank gij4886469[emb]CAB43385.1] - (AL050284) hypothetical protein [Homo sapiens]				264593
2685	88054289 (5389, 5370)	Novel Protein sim. GBank gij3342729 (AC005331) - R31341.2 [Homo sapiens]		UNCLASSIFIED		
2686	87628690 (5371, 5372)	Novel Protein sim. GBank gij4650844[db]BAA77027.1] - (AB026190) Keich motif containing protein [Homo sapiens]	Contains protein domain (PF01344) - struct Keich motif			264259, 29331822, 60432289, 29331827, 29331830, 284909, 284512, 264596, 264769, 284534, 264555, 264556, 264557, 264558, 60170394, 264559, 264486
2687	87698183 (5373, 5374)	Novel Protein sim. GBank gij5281314[gb]AAD41475.1[AF13312] - (AF133123) transcription factor IIC102 [Homo sapiens]	Contains protein domain (PF00515) - transcriptifactor TPR Domain			18108394, 18108396, 22278988, 35698288, 22278997, 29331826, 29331828, 68712502, 21906754, 265011, 264760, 264761, 284763, 284689, 21908765, 35698423, 264559, 18108385, 264583
2688	79959584 (5375, 5376)					284908, 264760
2689	94122440 (5377, 5378)	Novel Protein sim. GBank gij3880020[emb]CAA97339] - (Z73098) Similarity to yeast hypothetical protein (Swiss Prot accession number Q09895); cDNA EST EMBL:D72982 comes from this gene; cDNA EST EMBL:D75947 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00560) - Leucine Rich Repeat			22278997, 22278988, 22278999, 29331824, 35696052, 284908, 264808, 56182435, 264512, 264910, 265008, 60433438, 21906754, 18108351, 264682, 264683, 264767, 21908765, 21906766, 21906768, 33657023, 33657182, 27486262, 27486264, 33657349, 18108370, 18108372, 18108374, 35696423, 35695855, 18108385, 22279002
2690	88003055 (5379, 5380)	Novel Protein sim. GBank gij2477513 (AC002398) - F25965_3 [Homo sapiens]	Contains protein domain (PF00620) - struct RhoGAP domain			
2691	91219241 (5381, 5382)	Novel Protein sim. GBank gij4107276[emb]CAA67130] - (X98508) acetyl-CoA synthetase [Solanum tuberosum]	synthase			65274572, 56182575, 264259, 29331822, 29331824, 29331826, 29331828, 264112, 265009, 55812038, 264596, 33109954, 265017, 264448, 264288, 264369, 264684, 21906769, 60170615, 60431528, 55810764, 264634, 264636, 264556, 264637, 22279002, 264584, 264586
2692	94111914 (5383, 5384)	Novel Protein sim. GBank gij3513303 (AC005594) - R26984_1 [Homo sapiens]	peptidase			
2693	20438807 (5385, 5386)			UNCLASSIFIED		264592
2694	94111918 (5387, 5388)	Novel Protein sim. GBank gij3122400[sp]Q35682[MUG_MOUSE - MYELOID UPREGULATED PROTEIN	Prolyl oligopeptidase family			264559
2695	95345513 (5389, 5390)	Novel Protein sim. GBank gij4972740[gb]AAD34765.1] - (AF132177) unknown [Drosophila melanogaster]	collagen			35696286, 56994075, 22278989, 264259, 35696052, 29331830, 265011, 264288, 56181562, 264690, 264692, 33657023, 27486262, 263976, 18108376, 35698423, 35695855, 60170394, 83373044, 56526486, 22279000, 22279002, 264568
2696	87874040 (5391, 5392)	Novel Protein sim. GBank gij728831[sp]P39189[ALU1_HUMAN - IIII ALU SUBFAMILY J WARNING ENTRY III]	synthase			264594, 21906768, 18108370, 18108372

2697	91638472 (5393, 5394)	Novel Protein sim. GBank gi 5689473 dbj BAA33020.1  - (AB028991) KIAA1068 protein [Homo sapiens]		UNCLASSIFIED	35696286, 264259, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 265008, 265007, 264512, 265009, 60170831, 60433356, 264595, 264758, 87188474, 265010, 265011, 87188559, 264601, 265017, 265018, 265019, 264761, 18108351, 264446, 264288, 264766, 264688, 264689, 21908765, 21908768, 265020, 265021, 60170815, 33657109, 18108376, 35696423, 35695855, 264555, 264558, 60170394, 264559, 18108387, 56526486, 87168518, 22279002, 264563, 264482
2698	84325881 (5395, 5396)	Novel Protein sim. GBank gi 841318 (J22818) - mutant sterol regulatory element binding protein-2 [Cricetus griseus]	Contains protein domain (PF00412) - LIM domain containing proteins	transcript factor	22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 60432289, 29331828, 29331827, 33656970, 264906, 29331830, 264909, 52644045, 264910, 60433356, 33657402, 33109954, 265017, 265018, 265019, 264288, 21908765, 21908768, 21908767, 21908769, 29148629, 35695917, 265021, 265022, 52644150, 33657023, 33657182, 27486261, 35696423, 65274791, 264638, 60432113, 22279000
2699	87780650 (5397, 5398)			UNCLASSIFIED	264768, 18108357, 264690, 264691
2700	94139836 (5399, 5400)	Novel Protein sim. GBank gi 5174395 ref NP_006006.1 pB120 - Brain protein 120		UNCLASSIFIED	60424269, 56182435, 60432229, 60433438, 55811386, 265017, 55811150, 264448, 55818562, 55811957, 264693, 33657109, 60431528, 264629, 55810764, 55811576, 65274791, 60431850, 60432113
2701	94148584 (5401, 5402)	Novel Protein sim. GBank gi 1572801 (U70854) - F38A5.1 gene product [Caenorhabditis elegans]			18108394, 52645158, 35696286, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 35696052, 29331828, 29146499, 265006, 60433356, 33657402, 60433438, 264595, 33657084, 18108351, 264288, 264769, 18108359, 21908768, 35695917, 33657023, 27486261, 18108374, 18108379, 35696423, 18108382, 83373044, 18108384, 18108388, 60432113, 22279000
2702	57285368 (5403, 5404)	Novel Protein sim. GBank gi 2605967 (AF030027) - 24 [Equine herpesvirus 4]		UNCLASSIFIED	29331828, 264512, 264555, 264556, 264557, 264558, 264559
2703	87649514 (5405, 5406)	Novel Protein sim. GBank gi 5689399 dbj BAA82983.1  - (AB028954) KIAA1031 protein [Homo sapiens]	Contains protein domain (PF00642) - Zinc finger C-x8-C-x5-C-x3-H type (and similar).		60432289, 265007, 21908765, 21906788, 265021, 264563

2704	87646515 (5407, 5408)	Novel Protein sim. GBank gj 4335694 gb AAB53294  - (AF008594) Implantation-associated protein [Rattus norvegicus]			264488, 22278995, 22278998, 29331828, 28148489, 284805, 284808, 284807, 52644045, 264511, 33857402, 264600, 264602, 265017, 264605, 264781, 18108351, 264764, 264687, 264759, 265021, 264691, 264692, 18108382, 264693, 18108370, 18108374, 264634, 264635
2705	87771745 (5408, 5410)				264488, 264509, 264511, 264512, 264910, 264593, 87168474, 264604, 264286, 264687, 264769, 264638, 264588, 264486
2706	94326789 (5411, 5412)	Novel Protein sim. GBank gj 3255952 emb CAA16821.1  - (AL021728) /prediction=(method: [match=(desc: [Drosophila melanogaster])	UNCLASSIFIED		264488, 52646842, 65274572, 22278994, 56994075, 22278997, 264259, 29331824, 29331825, 29331826, 29331828, 33856970, 264907, 264908, 264909, 52644045, 56182435, 265006, 265007, 60433438, 55812038, 21906754, 52644298, 265010, 264601, 265017, 265019, 264681, 264448, 264682, 264286, 264688, 264687, 264688, 21906766, 21906769, 55811957, 35695317, 265020, 265021, 60170815, 264690, 264691, 33857023, 264692, 264693, 65274620, 27486284, 263972, 18108374, 18108377, 264635, 264636, 264556, 60170394, 83373044, 65274727, 87168518, 22278000
2707	86099839 (5413, 5414)	Novel Protein sim. GBank gj 3417294 (AC004381) - Unknown gene product [Homo sapiens]			22278996, 22278998, 56182435, 21906754, 87168559, 265017, 264448, 52645129
2708	91011351 (5415, 5416)	Novel Protein sim. GBank gj 545790 bbs 147178 - DARPP-32=dopamine and cAMP-regulated phosphoprotein [human, brain, Peptide, 204 aa]	UNCLASSIFIED		65274572, 264259, 29331822, 29331825, 60432289, 29331826, 29331827, 29331828, 264909, 264510, 265007, 264910, 60433356, 60433438, 33109954, 265010, 265011, 264389, 264288, 264785, 264693, 264585
2709	94853988 (5417, 5418)	Novel Protein sim. GBank gj 3169705 (AC004780) - F17127_1 [Homo sapiens]	UNCLASSIFIED		29331822, 18108370, 18108374, 83373044
2710	87627979 (5419, 5420)	Novel Protein sim. GBank gj 446831 emb CAB37892  - (AL031432) dJ46SN24.2.1 (PUTATIVE novel protein) (isoform 1) [Homo sapiens]			29331824, 264759, 264693, 18108382, 18108388

2711	94111920 (5421, 5422)	Novel Protein sim. GBank gi 3122400 sp O35682 MUG_MOUSE - MYELOID UPREGULATED PROTEIN		UNCLASSIFIED	264488, 264687, 52645156, 264768, 21906764, 21906765, 21906767, 21906768, 21906769, 55811957, 56994075, 22278997, 22278998, 265020, 265021, 264690, 264259, 264691, 264692, 33657023, 29331822, 29331824, 60424269, 29331826, 33657182, 29331827, 27488262, 33657348, 264508, 264905, 264907, 60431528, 264908, 264909, 55810764, 35696423, 65274791, 35695855, 265007, 264910, 60431850, 60432229, 264557, 264558, 55812038, 33108954, 18108385, 21906754, 33657084, 87168518, 87168474, 87168559, 60432113, 265017, 22279000, 265018, 265019, 22279002, 264760, 55811150, 264681, 18108351, 264565, 264764, 264566, 264288, 264766 264488, 35696286, 22278998, 264259, 29331824, 60432269, 35696052, 264508, 264908, 66712502, 52644045, 265006, 60432229, 33657402, 60433356, 265010, 265019, 18108351, 264681, 264288, 264685, 21906765, 21906766, 21906768, 21906769, 55811957, 35695917, 285020, 265021, 60170815, 264691, 264592, 33657023, 33657109, 33657182, 33657349, 18108370, 18108374, 35696423, 35695855, 264555, 52644332, 56182323, 87168518, 60432113 66714117, 264906, 264563
2712	94312071 (5423, 5424)	Novel Protein sim. GBank gi 5081315 gb AAD39343.1 AF07660 - (AF076607) prediabetic NOD sera-reactive autoantigen [Mus musculus]	Contains protein domain (PF00515) - TPR Domain		
2713	88003064 (5425, 5426)	Novel Protein sim. GBank gi 5081315 gb AAD39343.1 AF07660 - (AF076607) prediabetic NOD sera-reactive autoantigen [Mus musculus]		UNCLASSIFIED	
2714	13528218 (5427, 5428)	Novel Protein sim. GBank gi 5081315 gb AAD39343.1 AF07660 - (AF076607) prediabetic NOD sera-reactive autoantigen [Mus musculus]		UNCLASSIFIED	264636
2715	94122454 (5429, 5430)	Novel Protein sim. GBank gi 4321868 gb AAD15897 - (AF067430) Smarcat1-related protein [Mus musculus]		UNCLASSIFIED	264508, 264905, 264907, 264908, 264909, 264910, 264758, 265011, 264760, 264764, 264268, 264766, 264768, 264636
2716	88003068 (5431, 5432)	Novel Protein sim. GBank gi 2477513 (AC002398) - F25965_3 [Homo sapiens]		glycoprotein	264091, 264259, 29331822, 66714117
2717	80077461 (5433, 5434)	Novel Protein sim. GBank gi 3327046 dbj BAA31591  - (AB014516) KIAA0616 protein [Homo sapiens]			264908, 264389, 264893, 264556, 264563 264593, 264558
2718	79604062 (5435, 5436)	Novel Protein sim. GBank gi 746495 (U23515) - weakly similar to gastrula zinc finger protein [Caenorhabditis elegans]			264693
2719	88180423 (5437, 5438)	Novel Protein sim. GBank gi 746495 (U23515) - weakly similar to gastrula zinc finger protein [Caenorhabditis elegans]		UNCLASSIFIED	29331822, 87168559, 265019, 265021, 52644150, 264691

2720	95066242 (5439, 5440)	Novel Protein sim. GBank gi 1335873 (U6690) - ATP-dependent RNA helicase [Mus musculus]	Contains protein domain (PF00270) DEAD/DEAH box helicase	helicase	18108374, 60424178, 264489, 56182435, 21906765, 21806766, 35698423, 22278997, 265020, 265022, 265005, 265008, 264092, 264636, 60432228, 264691, 264692, 33657023, 264693, 33657402, 83373044, 29331824, 18108366, 60424269, 29331826, 18108385, 52645129, 21906784, 35698052, 29331828, 87168474, 264100, 265010, 265011, 265018, 22279002, 264905, 264482, 284563, 264908, 18108351, 284681, 18108370, 29331830, 264908, 68712502, 52644045, 264909, 264828, 18108354, 22278995, 35696286, 264259, 29331822, 29331824, 66714117, 29331826, 264906, 60433438, 265017, 18108351, 264448, 284288, 264769, 21906786, 265021, 33657109, 263868, 60431528, 264629, 55811576, 65274791, 35695855, 264631, 264637, 60170394, 56182323, 22279000, 35696286, 22278999, 21906754, 265017, 264762, 264288, 21906765, 21906767, 21906788, 35695917, 18108382, 27486282, 35695855, 264558, 264559
2721	95345523 (5441, 5442)	Novel Protein sim. GBank gi 4929653 (gb AAD34092.1) AF15185 - (AF151855) CGI-97 protein [Homo sapiens]	Contains protein domain (PF01172) - Uncharacterized protein family UPF0023		
2722	91638607 (5443, 5444)	Novel Protein sim. GBank gi 3212997 (gb AAC23434.1) - (AC004987) match to ESTs AA667999 (NID:g2626700), AA165465 (NID:g1741481), Z45871 (NID:g575105), and T84026 (NID:g12314); similar to various Irf-like proteins including: AF040654 (PID:g2746883), D13644 (PID:g2104571), AL02114...	Contains protein domain (PF00566) - TBC domain	oncogene	
2723	87387732 (5445, 5446)			UNCLASSIFIED	264508, 264509, 264906, 264809, 264910, 55812038, 264765, 264687, 264628, 264636, 264485
2724	87639563 (5447, 5448)	Novel Protein sim. GBank gi 4680681 (gb AAD27730.1) AF13295 - (AF132955) CGI-21 protein [Homo sapiens]		ubiquitin	18108396, 22278998, 20281088, 29331824, 29331826, 60432289, 29331828, 60170831, 60432229, 60433438, 18108351, 264682, 21906766, 21906767, 21906769, 35695917, 33657023, 33657109, 18108372, 18108374, 35695855, 22279000, 22279002
2725	94853891 (5448, 5450)	Novel Protein sim. GBank gi 3169705 (AC004780) - F17127_1 [Homo sapiens]		UNCLASSIFIED	284488, 52644507, 264259, 29331827, 21906754, 285011, 18108351, 284448, 284288, 264685, 264689, 35695917, 265020, 33657182, 27486261, 18108370, 18108374, 35698423, 18108385, 22279000
2726	86880599 (5451, 5452)	Novel Protein sim. GBank gi 3342738 (AC005328) - R26680_1, partial CDS [Homo sapiens]		MHC	264488, 264828, 264685

2727	91010470 (5453, 5454)	Novel Protein sim. GBank gjl731267/sp139219/LUA_ECOLI - RIBOSOMAL LARGE SUBUNIT PSEUDOURIDINE SYNTHASE A (PSEUDOURIDYLATE SYNTHASE) (URACIL HYDROLYASE)	Contains protein domain (PF00849) - RNA pseudouridyate synthase	- synthase	65274572, 56182575, 22278994, 56984075, 22278996, 22278997, 22278999, 60432049, 29331822, 29331824, 66714117, 29331828, 29331827, 35896052, 29331828, 33656970, 264509, 66712502, 264910, 33657402, 60433438, 264758, 55812038, 21908754, 33657084, 55811386, 265018, 265019, 264767, 21906765, 21906767, 21906769, 55811957, 35895917, 52644150, 33657023, 33657109, 33657182, 27486261, 27486282, 27486285, 33657349, 55811576, 35898423, 35898555, 264630, 60431850, 264638, 56182323, 87188518, 60432113, 22278000, 264564, 264565
2728	94126022 (5455, 5456)	Novel Protein sim. GBank gjl3880433/lemb/CAA91399 - (Z66521) similar to mitochondrial RNA splicing MSR4 like protein, cDNA EST EMBL:C09217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	18108394, 56182435, 21906767, 55811957, 35898555, 265021, 264690, 264596, 264259, 264557, 29331822, 264559, 264448, 264288
2729	94126024 (5457, 5458)	Novel Protein sim. GBank gjl2408095/lemb/CAB18300 - (Z99168) putative RNA splicing protein [Schizosaccharomyces pombe]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	65274572, 264259, 60432289, 66712502, 56182435, 264448, 264288, 264399, 55811957, 265021, 264557, 60432113
2730	94126026 (5459, 5460)	Novel Protein sim. GBank gjl3880433/lemb/CAA91399 - (Z66521) similar to mitochondrial RNA splicing MSR4 like protein, cDNA EST EMBL:C09217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	264887, 29331824, 29331826, 35896052, 264107, 56182435, 265008, 265009, 264592, 60431735, 265011, 264601, 265017, 18108351, 264288, 29148627, 55811957, 265021, 264690, 18108368, 18108374, 264557, 264558, 264559, 18108387, 56526488, 264566, 264466
2731	87723022 (5461, 5462)	Novel Protein sim. GBank gjl1723239/sp101661/VAUB_SCHPO - HYPOTHETICAL 35.7 KD PROTEIN C26A3.11 IN CHROMOSOME 1	Contains protein domain (PF00795) - Carbon-nitrogen hydrolase		264259, 35896052, 265006, 264758, 264762, 264448, 264288, 29148627, 21908769, 87168518, 22279002
2732	94126028 (5463, 5464)	Novel Protein sim. GBank gjl3880433/lemb/CAA91399 - (Z66521) similar to mitochondrial RNA splicing MSR4 like protein, cDNA EST EMBL:C09217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	264887, 264488, 18108358, 56182435, 264888, 35898423, 55811957, 265021, 265006, 265008, 264910, 265009, 264690, 264555, 264259, 264556, 264557, 264598, 264559, 18108383, 33657109, 87168518, 265010, 264601, 60432113, 265017, 264905, 264448, 263972, 264369, 264567
2733	87363060 (5465, 5466)			UNCLASSIFIED	29331825, 264509, 264809
2734	94140286 (5467, 5468)	Novel Protein sim. GBank gjl4519621/dbj/BA475670.1 - (AB017614) OASIS protein [Mus musculus]	Contains protein domain (PF00170) - bZIP transcription factor	transcription factor	60424179, 52844507, 56182575, 264259, 29331828, 264807, 264510, 264910, 60433356, 265019, 55811150, 264681, 264763, 264687, 33657182, 18108370, 60431528, 60431850, 56182323, 83373044

2735	87712338 (5489, 5470)	Novel Protein sim. GBank gij3850569 (AC005278) - ESTs gb T21278, gb T45403, and gb AA586113 come from this gene. [Arabidopsis thaliana]		glycoprotein	22278996, 60432289, 29331827, 26146498, 264108, 264909, 264112, 33657402, 87168474, 265017, 264762, 264448, 264764, 264684, 21906765, 264693, 33657109, 263976, 264636, 264638, 264557, 22279000, 22279002, 264587
2736	80247855 (5471, 5472)			UNCLASSIFIED	264905, 264628, 264628, 263978, 264632, 264564
2737	87604526 (5473, 5474)				264690
2738	85731808 (5475, 5476)	Novel Protein sim. GBank gij2558501[dbj AA22896] - (D63850) hepaloma-derived growth factor [Mus musculus]			264486, 265008, 264768, 264691
2739	94318834 (5477, 5478)	Novel Protein sim. GBank gij5420387[emb CA846679.1] - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264684, 83373044, 264566
2740	94148762 (5479, 5480)	Novel Protein sim. GBank gij3417386[emb CAA75495] - (Y15197) microtubule-associated protein, MAP-115 [Mus musculus]		UNCLASSIFIED	264488, 56182575, 22278995, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 29146498, 29331830, 265006, 265007, 265008, 60432229, 33657402, 55812038, 87168474, 265010, 265011, 265017, 265018, 265019, 264605, 264681, 264288, 264369, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265022, 264691, 264682, 33657109, 18108370, 18108374, 55810764, 35695855, 264634, 60431850, 264639, 56182323, 18108382, 18108385, 65274727, 22279002, 264584
2741	88047518 (5481, 5482)	Novel Protein sim. GBank gij3242764 (AC005154) - similar to protein U28928 (P1Dg861306) [Homo sapiens]		UNCLASSIFIED	22278996, 52644045, 52644229, 21906768, 21906769, 265020, 60170615, 264691
2742	87648644 (5483, 5484)	Novel Protein sim. GBank gij4758412[ref NP_004472.1]pGALN - UDP-N-acetyl-alpha-D-galactosamine polypeptide N-acetyl-galactosaminyltransferase 2 (Galinac-T2)	Contains protein domain (PF00652) - Similarity to lectin domain of ricin beta-chain, 3 copies.	transferase	264259, 264905, 264758, 55812038, 264369, 29148627
2743	87627991 (5485, 5486)	Novel Protein sim. GBank gij446831[emb CA837982] - (AL031432) dJ465N24.2.1 (PUTATIVE novel protein) (isoform 1) [Homo sapiens]		UNCLASSIFIED	35696286, 264259, 264908, 264908, 265008, 60433438, 265017, 18108351, 264448, 264764, 264288, 21906765, 21906767, 264690, 264691, 264893, 263999, 263971, 35695855, 264637, 264558, 18108382, 60432113
2744	94126030 (5487, 5488)	Novel Protein sim. GBank gij3880433[emb CAA91399] - (Z66921) similar to mitochondrial RNA splicing MSR4 like protein: cDNA EST EMBL:CO9217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	kinase	18108374, 264488, 56182435, 21906765, 35696423, 35695917, 35695855, 265020, 265021, 264511, 265009, 264490, 264556, 264258, 264557, 56182323, 264558, 294559, 18108383, 29331824, 18108385, 33657109, 29331826, 21906754, 29331827, 29331828, 33657349, 87168518, 265018, 264905, 264482, 264448, 264488, 264368, 264288

2745	87740125 (5489, 5490)	Novel Protein sim. GBank gij4405795gijAAD198261 - (AF038963) RNA helicase [Homo sapiens]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	35696286, 284509, 284905, 284907, 284908, 284909, 284510, 284512, 265008, 264758, 284601, 265017, 284604, 284763, 284288, 284686, 284759, 284693, 35696423, 35695855, 284634, 284636, 284563, 284564, 264565
2746	95418601 (5491, 5492)	Novel Protein sim. GBank gij4758738refNP_004680.1pMTA1 - metastasis associated 1	Contains protein domain (PF00320) - GATA zinc finger	UNCLASSIFIED	22278996, 22278998, 22278999, 28331822, 28331826, 29331827, 35696052, 28331828, 284905, 284906, 284907, 284908, 284909, 52644045, 285006, 60170831, 284598, 55812038, 285018, 284683, 284288, 21906765, 21906767, 21906768, 21906769, 265020, 264690, 33657023, 284693, 33857109, 18108388, 18108374, 284558, 18108385, 22278000, 284563
2747	94112677 (5483, 5494)	Novel Protein sim. GBank gij4557803refNP_000262.1pNPC1 - Niemann-Pick disease, type C1		glycoprotein	264569, 52644507, 18108394, 22278995, 35696286, 22278997, 22278999, 52645080, 29331824, 56182181, 29331826, 29331827, 35696052, 284907, 284908, 284909, 265009, 33109854, 55811386, 87168474, 265010, 87168559, 284603, 265019, 284760, 284688, 284768, 21908769, 35695917, 60170615, 284692, 33657023, 52645128, 27488284, 60431528, 18108374, 35696423, 35695855, 264558, 56182323, 18108385, 264482
2748	91214883 (5485, 5486)	Novel Protein sim. GBank gij4191272emb(CAA03984) - (AJ012295) epaG protein [Rhizobium elii]	Contains protein domain (PF00646) - F-box domain		65274572, 29331828, 264112, 264511, 265019, 284760, 284767, 284768, 284769, 21908768, 21908769, 285020, 27488282, 56526488, 87168518, 22278000
2749	87346307 (5497, 5498)				264259, 284908, 284510, 265008, 265009, 264760, 284389, 284768, 284563
2750	87335344 (5499, 5500)	Novel Protein sim. GBank gij1872498 (U74297) - PIUS [Cryptolagus cuniculus]		UNCLASSIFIED	264488, 52644507, 18108396, 56994075, 264259, 29331825, 29331826, 29331827, 29331828, 284508, 265009, 284910, 284591, 284595, 33857084, 285011, 285019, 18108351, 284288, 284688, 284768, 284689, 55811957, 284693, 27488284, 18108370, 18108374, 284558, 18108385, 264482, 264563
2751	87057465 (5501, 5502)			UNCLASSIFIED	29331822, 29331824, 265017, 33657023
2752	88062875 (5503, 5504)	Novel Protein sim. GBank gij3041859 (AC004534) - OG-2 homeodomain protein-like, similar to U65067 (PID:g1575528) [Homo sapiens]	Contains protein domain (PF00046) - Homeobox domain	homeobox	



2753	94138972 (5505, 5506)	Novel Protein sim. GBank gi 33651648 (AF096301) - neural F-box protein NFB42 [Rattus norvegicus]	Contains protein domain (PF00846) - F-box domain.	UNCLASSIFIED	56182575, 56994075, 22278998, 29331822, 29331824, 29331825, 29331828, 265007, 264593, 55812038, 33108954, 18108351, 264288, 56181562, 21906768, 21906768, 265021, 2646893, 18108374, 85274791, 264632, 58182323, 22279002, 264563, 264587
2754	94115513 (5507, 5508)	Novel Protein sim. GBank gi 535428 (U13738) - calmodulin-like protein [Pisum sativum]	Contains protein domain (PF00036) - EF hand	struct	22278999, 66714117, 29331827, 35696052, 29331828, 264906, 264908, 264909, 265011, 265017, 265018, 265019, 264288, 21906765, 21906767, 265022, 33657023, 264893, 56182323, 18108382, 22279000
2755	88001472 (5508, 5510)	Novel Protein sim. GBank gi 2996653 (AC004510) - R30385_2 [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	
2756	11465908 (5511, 5512)	Novel Protein sim. GBank gi 173539 (U30473) - putative src-like adapter protein; non-catalytic src-like adapter protein containing SH3 and SH2 domains; homolog of mouse SLAP; Method: conceptual translation supplied by author [Homo sapiens]	Contains protein domain (PF00017) - Src homology domain 2	UNCLASSIFIED	264594
2757	95381590 (5513, 5514)	Novel Protein sim. GBank gi 3608372 (AF053758) - brain specific cortactin-binding protein CBP90 [Rattus norvegicus]		eph	85658542, 264693
2758	79637846 (5515, 5516)	Novel Protein sim. GBank gi 3608372 (AF053758) - brain specific cortactin-binding protein CBP90 [Rattus norvegicus]		struct	264693
2759	91005312 (5517, 5518)	Novel Protein sim. GBank gi 2072200 (U94863) - p40 [Borna disease virus]			65274572, 35696286, 66714117, 29331828, 264508, 56182435, 21906754, 55811957, 264629, 264638, 56182323, 22279002, 264908
2760	79824798 (5519, 5520)	Novel Protein sim. GBank gi 4914573 (lemb) CAB43885.1 - (AL050390) hypothetical protein [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	UNCLASSIFIED	22278998, 29331822, 29331830, 265010, 265019, 264288, 21906765, 21906768, 21906769, 265020, 56182323, 22279002, 264563
2761	87639597 (5521, 5522)	Novel Protein sim. GBank gi 3136150 (AF050182) - PERIOD 3 [Mus musculus]	Contains protein domain (PF00889) - PAS domain	struct	18108394, 22278998, 264806, 264909, 265006, 265007, 264757, 265010, 265011, 265017, 265018, 18108351, 264448, 264683, 264686, 264768, 265020, 265021, 265022, 264691, 18108382, 264693, 18108385, 33657109, 18108388, 18108370, 18108381, 18108382, 18108384, 18108388, 87168518, 264369, 35696423
2762	87592699 (5523, 5524)	Novel Protein sim. GBank gi 3136150 (AF050182) - PERIOD 3 [Mus musculus]	Contains protein domain (PF01352) - KRAB box	transcriptfactor	
2763	87539968 (5525, 5526)	Novel Protein sim. GBank gi 3511122 (AF060503) - zinc finger protein [Homo sapiens]	Contains protein domain (PF00370) - FCGY family of carbohydrate kinases		52646842, 264259, 29331822, 29331825, 29331828, 29331828, 33656970, 264905, 264907, 29331830, 265006, 265009, 21906754, 265019, 264448, 21906769, 27488292, 56182323, 56526488, 87168518, 264487
2764	94305140 (5527, 5528)	Novel Protein sim. GBank gi 2905843 (AF045244) - ribitol kinase [Klebsiella pneumoniae]			

2765	94315105 (5529, 5530)	Novel Protein sim. GBank gij4688672[embjCAA17689.2] - (AL022016) /prediction=(method:; /prediction=(method:; /match=(desc: [Drosophila melanogaster])		264488, 65274572, 22278995, 35696286, 22278998, 22278999, 264259, 29331822, 29331824, 35696052, 264508, 264509, 264905, 264908, 264907, 264908, 264909, 264512, 264910, 265009, 264592, 264595, 264758, 55812038, 33109954, 265010, 87168559, 264600, 265018, 264760, 264761, 264762, 264763, 264448, 264764, 264288, 264768, 264767, 264768, 56181562, 21908764, 21908765, 21908768, 35695917, 265021, 264691, 264692, 33657023, 33657109, 264628, 18108370, 264629, 18108374, 55811578, 35698423, 35695855, 264631, 264634, 264635, 264636, 264637, 264638, 264558, 264639, 83373044, 87168518, 22279000, 22279002, 264563, 264482
2766	94315109 (5531, 5532)	Novel Protein sim. GBank gij5441611[embjCAB46854.1] - (AJ388555) hypothetical protein (Canis familiaris)	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	264488, 65274572, 60432289, 264907, 264909, 264511, 264512, 60433356, 264288, 264685, 264688, 35695917, 265022, 264693, 264628, 65274791, 264635, 264555, 264556, 264557, 264636, 264558, 264559, 83373044, 60432113
2767	80204297 (5533, 5534)	Novel Protein sim. GBank gij1079451[prjA55463 - tropomodulin, skeletal muscle - chicken]	struct	264112, 263974, 264558
2768	94322238 (5535, 5536)	Novel Protein sim. GBank gij5441322[embjCAB46721.1] - (AL031427) dJ167A19.1 (novel protein) [Homo sapiens]	UNCLASSIFIED	264488, 263994, 264489, 35696286, 22278998, 22278999, 264259, 29331822, 60432289, 29331826, 35696052, 264508, 264509, 264905, 264906, 264907, 29331830, 264908, 264909, 264510, 264511, 264512, 265008, 264910, 265009, 264591, 264592, 264593, 33657402, 264594, 264595, 264757, 264596, 265011, 264604, 265018, 264760, 264681, 18108351, 264764, 264288, 264766, 264686, 264768, 18108357, 264769, 264689, 21908766, 21908768, 21908769, 35695917, 265021, 265022, 33657023, 33657109, 264628, 264629, 18108374, 35698423, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264555, 264638, 264639, 83373044, 264563, 264564, 264565, 264566, 264488, 264567

2769	95311088 (5537, 5538)	Novel Protein sim. GBank gij5419859jembj(CAB46376.1) - (AL096725) hypothetical protein [Homo sapiens]		tubulin	264488, 56182575, 264259, 29331822, 29331824, 68714117, 29331825, 60432289, 29331828, 264908, 29331830, 56182435, 264592, 33657402, 264448, 264389, 264288, 60170815, 264891, 33657023, 264692, 33657108, 18108374, 55811576, 264834, 264636, 56182323, 83373044, 60432113
2770	87730182 (5539, 5540)	Novel Protein sim. GBank gij5701865jembj(CAB52157.1) - (AL109736) WD repeat protein [Schizosaccharomyces pombe]		kinase	18108398, 22278995, 22278996, 22278999, 264105, 265006, 265018, 18108351, 264687, 21908765, 18108364, 264629, 18108374, 264631, 18108385, 18108388
2771	86084071 (5541, 5542)	Novel Protein sim. GBank gij3093433 (AC004125) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	264259, 29331822, 29331824, 29331825, 264389
2772	95357308 (5543, 5544)	Novel Protein sim. GBank gij4895531ref[INP_005465.1]pNY C - histone deacetylase 5		histone	264259, 29331822, 29331824, 68714117, 60432289, 29331827, 264905, 285008, 264592, 55812038, 65274444, 264768, 21908769, 33657109, 263978, 264595, 264638, 264557, 83373044, 264583, 264584, 264486
2773	94138984 (5545, 5546)	Novel Protein sim. GBank gij3288888 (AC005253) - R26445_1 [Homo sapiens]		UNCLASSIFIED	18108374, 264686, 264687, 263976, 56182435, 264688, 55810784, 21908766, 35896423, 55811578, 65274791, 56181888, 55811957, 35895855, 264110, 265021, 264112, 265022, 285008, 265008, 264092, 264084, 60431850, 264637, 264691, 264259, 264692, 263981, 264594, 60433356, 56182323, 264693, 264757, 56182181, 55812038, 29331825, 60424269, 18108385, 29331826, 29331827, 27488261, 29331828, 35896052, 55811386, 264107, 60432113, 265017, 55811150, 18108351, 264681, 264908, 18108370, 264484, 264682, 20281069, 284448, 68712502, 264683, 264764, 264288, 264684, 264758, 263974
2774	87818908 (5547, 5548)	Novel Protein sim. GBank gij465852spjP34388)YLS3_CAEEL - HYPOTHETICAL 70.7 KO PROTEIN P09G8.3 IN CHROMOSOME III		ribosomalprot	22278995, 35896288, 22278996, 22278999, 264259, 29331826, 60432289, 35896052, 264112, 33657402, 21908754, 87168559, 265017, 265018, 264288, 21908766, 21908767, 21908769, 35895917, 265020, 265021, 33657023, 33657109, 18108370, 263978, 35896423, 35895855, 87168518, 22278000, 264482

2775	95307887 (5548, 5550)	Novel Protein sim. GBank gi4689132 gb AAD27775.1 AF07704 - (AF07704) 30S ribosomal protein S7 homolog [Homo sapiens]	Contains protein domain (PF00177) - Ribosomal protein S7p/Ssa	264488, 22278995, 56994075, 22278996, 35696286, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 80432289, 29331827, 29331828, 35696052, 265007, 21906754, 265017, 265019, 264448, 264682, 264369, 264288, 18108354, 52644229, 264688, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265021, 265022, 60170815, 264691, 18108370, 35696423, 65274791, 35695855, 264634, 60431850, 60170394, 56182323, 264558, 18108388, 22279000, 264563, 264565, 264486, 264567
2776	87701557 (5551, 5552)			56182575, 22278998, 22278999, 264259, 29331822, 29331824, 264908, 29331830, 264510, 33657402, 21906754, 55811388, 265017, 265019, 264448, 264288, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 265022, 60170615, 55810764, 55811576, 264555, 56526486, 22279000
2777	79818728 (5553, 5554)		UNCLASSIFIED	264807, 264766
2778	82112411 (5555, 5556)		UNCLASSIFIED	264807, 264593, 264760, 264628
2779	87649729 (5557, 5558)	Novel Protein sim. GBank gi4680711 gb AAD27745.1 AF13297 - (AF13297) CGI-36 protein [Homo sapiens]	UNCLASSIFIED	22278997, 264259, 29331824, 66714117, 35696052, 265006, 264512, 264448, 264288, 29148827, 18108364, 20281149, 18108370, 264629
2780	94679397 (5559, 5560)	Novel Protein sim. GBank gi4758524 ref NP_004825.1 pH-GK - HPK/GCK-like kinase	Contains protein domain (PF00780) - CNH domain	29146499, 65274791, 264634, 264639
2781	91220057 (5561, 5562)	Novel Protein sim. GBank gi468352 gb AAD21222 - (AF069502) ubiquitin specific protease UBP43 [Mus musculus]	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	60424179, 29331824, 60424269, 66714117, 29331828, 56182435, 87188474, 265017, 264764, 56181562, 21906765, 21906766, 21906768, 35695917, 265020, 33657023, 35695855, 56182323, 87188518
2782	94233146 (5563, 5564)	Novel Protein sim. GBank gi4505013 ref NP_002310.1 pLRN - leucine-rich neuronal protein	Contains protein domain (PF00560) - Leucine Rich Repeat	65274572, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331828, 60432289, 29331828, 264905, 264807, 264908, 264909, 52644045, 265009, 265017, 265018, 264604, 265019, 264760, 264683, 264288, 264766, 264685, 264688, 264768, 52644229, 264689, 21906768, 265020, 265021, 264691, 18108382, 264692, 33657023, 264693, 33657109, 33657349, 18108370, 264628, 263978, 35695855, 264557, 56182323, 83373044, 18108385

2783	60016628 (5585, 5566)	Novel Protein sim. GBank gi 728831 sp P39188 ALU1_HUMAN - III ALU SUBFAMILY J WARNING ENTRY IIII		lm7	264909, 264628, 263978, 263981
2784	87614360 (5587, 5568)				264259, 20331822, 29331824, 29331825, 264482
2785	88071930 (5589, 5570)	Novel Protein sim. GBank gi 2134933 p rf S58890 - collapsin response mediator protein - human	UNCLASSIFIED		22278999, 264908, 264758, 265018, 264789, 21806785, 21806788, 21806789, 265020, 264584
2786	87408542 (5571, 5572)	Novel Protein sim. GBank gi 2073564 (U80223) - eukaryotic inflation factor eIF-2 alpha kinase; DGCN2 [Drosophila melanogaster]	Contains protein domain (PF00069) Eukaryotic protein kinase domain	kinase	264905, 265017
2787	87901266 (5573, 5574)	Novel Protein sim. GBank gi 5174507 ref NP_006020.1 pMA1 - paraneoplastic neuronal antigen	UNCLASSIFIED		264768, 21906765, 21906788, 55811957, 22278999, 264093, 60170615, 264259, 29331822, 18108365, 29331824, 33657109, 29331827, 35696052, 264100, 264105, 264908, 263977, 55811576, 264635, 264637, 60433438, 60432113, 265017, 265019, 22279002, 55811150, 264359, 264288
2788	88080644 (5575, 5576)	Novel Protein sim. GBank gi 3252828 (AC004382) - Unknown gene product [Homo sapiens]	UNCLASSIFIED		
2789	85491275 (5577, 5578)	Novel Protein sim. GBank gi 2495729 sp Q92558 Y281_HUMAN - HYPOTHETICAL PROTEIN KIA0281 (HA8725)	UNCLASSIFIED		264103, 21906769, 264693
2790	87602784 (5579, 5580)	Novel Protein sim. GBank gi 5101772 emb CAB45135.1  - (AJ242976) p621 [Homo sapiens]	struct		264488, 264768, 264910, 264631, 264635, 264690, 264691, 264259, 264838, 29331824, 264508, 264509, 264905, 264583, 264908, 264628, 18108370, 264907, 264764, 264908, 264288, 264909
2791	88083195 (5581, 5582)	Novel Protein sim. GBank gi 2911266 (AC002550) - Unknown gene product [Homo sapiens]			21906784, 18108368
2792	95083783 (5583, 5584)	Novel Protein sim. GBank gi 2854163 gb AAC02581.1  - (AF045842) No definition line found [Caenorhabditis elegans]			22278995, 22278997, 264259, 29331822, 29331824, 29331825, 68714117, 29331826, 60432289, 29331827, 35696052, 264908, 68712502, 29331830, 284909, 60432229, 60433356, 60433438, 33109954, 285010, 265011, 265017, 265018, 265019, 264682, 264448, 264288, 264359, 264766, 52644229, 21906765, 21906768, 265020, 265021, 33657023, 263974, 18108374, 65274781, 35698856, 264636, 264558, 264558, 56182323, 83373044, 18108385, 36326486, 60432113, 22279000, 284587
2793	87425476 (5585, 5586)		UNCLASSIFIED		264259, 60432289, 66712502, 265009, 264638
2794	85794830 (5587, 5588)		UNCLASSIFIED		264689, 265022, 264691, 18108388, 264567

2795	95334888 (5589, 5590)	Novel Protein sim. GBank gi 5454148 ref NP_006348.1 pUBB2 - ubiquitin-conjugating enzyme E2E 3 (homologous to yeast UBC4/5)	Contains protein domain (PF00179) - Ubiquitin-conjugating enzyme	ubiquitin	55274572, 56182575, 35686288, 22278998, 22278999, 60432049, 264259, 28331822, 29331825, 66714117, 60432288, 29331828, 29331827, 35686052, 29331828, 264907, 66712502, 58182435, 284511, 265007, 264512, 264910, 60432229, 60433356, 60433438, 33109954, 85658542, 265018, 265019, 264288, 264686, 21906784, 21906765, 21906766, 21906768, 21908788, 55811957, 265020, 265021, 265022, 52644150, 33657023, 264893, 65274620, 33657109, 35596423, 55811576, 65274791, 56182323, 56526486, 60432113, 22279002, 264482, 264563, 264484, 264567
2798	94848857 (5591, 5592)	Novel Protein sim. GBank gi 4690657 gb AAD27715.1 AF13294 - (AF132840) CGI-08 protein [Homo sapiens]		UNCLASSIFIED	18108394, 65274572, 56182575, 56994075, 22278999, 264490, 60432049, 264259, 29331822, 29331824, 29331826, 35686052, 264509, 264908, 264907, 284908, 66712502, 56182435, 264510, 265008, 264512, 265007, 265008, 264910, 265009, 264591, 264592, 60433356, 264594, 264595, 55812038, 264596, 21906754, 60174639, 87168474, 265010, 265011, 265017, 265018, 265019, 55811150, 264762, 18108351, 264448, 264682, 264369, 264288, 264685, 264766, 264687, 56181582, 264769, 264689, 21908785, 21908766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 60170615, 52644150, 264692, 33657023, 18108362, 264693, 65274620, 33657109, 33657182, 27486265, 33657349, 18108374, 35696423, 65274791, 35695855, 264556, 264557, 56182323, 264558, 60170394, 83373044, 65274727, 87168518, 22279000, 264563, 264584, 264585, 264566, 264567
2797	95110790 (5593, 5594)	Novel Protein sim. GBank gi 4838557 gb AAD31040.11 - (AF143859) potassium channel modulatory factor DEB1-91 [Mus musculus]	Contains protein domain (PF00569) - Zinc finger present in dystrophin, CBP/p300	UNCLASSIFIED	56182575, 22278995, 22278996, 22278997, 29331827, 29146499, 264509, 264908, 56182435, 264757, 21906784, 265010, 265017, 265018, 284881, 264682, 264683, 264688, 21908765, 21906767, 21906768, 21908789, 29148629, 265020, 52644150, 264690, 33657182, 264629, 18108376, 56182323, 22279002, 264563
2798	86188005 (5595, 5596)	Novel Protein sim. GBank gi 2852645 (AF007160) - unknown [Homo sapiens]		UNCLASSIFIED	264488, 264480, 264259, 264448, 20281149, 20281152, 264556, 284557, 264558, 264559, 264483, 264486, 264567

2789	88080651 (5597, 5598)	Novel Protein sim. GBank gij3252825 (AC004382) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331828, 35696052, 29331828, 264909, 60433356, 33657402, 33109954, 87168474, 264448, 52644229, 21906765, 21906767, 21906768, 35695917, 265020, 265021, 265022, 52644150, 35695955, 264834, 60432113, 22279000
2800	88316481 (5589, 5600)	Novel Protein sim. GBank gij4240301 [gbjBAA74829.1] - (AB020713) KIAA0906 protein [Homo sapiens]		glycoprotein	264488, 56994075, 264259, 20281099, 29331825, 29331827, 264905, 56182435, 265008, 265011, 87168559, 265017, 265019, 264448, 264288, 264766, 264688, 60170815, 264691, 264692, 27488265, 264628, 264629, 264636, 264557, 264558, 264559, 87168518, 264584, 264586, 264587
2801	86068814 (5601, 5602)		Contains protein domain (PF00627) - UBA domain		265007, 264687
2802	88082477 (5603, 5604)	Novel Protein sim. GBank gij2337865 (AC002454) - organic cation transporter; 50% similarity to JC4884 (PID:g2143892) [Homo sapiens]	Contains protein domain (PF00083) - transport Sugar (and other) transporter		264448, 35695855
2803	79577445 (5605, 5606)	Novel Protein sim. GBank gij4559368 [g01AAD2029.1] AC006585	Contains protein domain (PF01585) - peptidase	UNCLASSIFIED	284639
2804	57111131 (5607, 5608)	Novel Protein sim. GBank gij4559368 [g01AAD2029.1] AC006585	G-patch domain		264586
2805	87398488 (5609, 5610)				264092, 264259, 29331822, 29331824, 29331827, 29331828, 264508, 265007, 265009, 21906754, 264389, 264288, 264632, 60170394, 264583, 264482
2806	87898951 (5611, 5612)	Novel Protein sim. GBank gij1168973 [p44403] CLPB_HAEN - CLPB PROTEIN		UNCLASSIFIED	22278995, 22278996, 22278999, 264259, 29331822, 29331824, 29331826, 29331827, 29146498, 265008, 265009, 60433438, 265017, 265018, 265019, 264448, 264288, 21906765, 21906767, 21906769, 28148629, 265022, 52644150, 56182323

2807	91720702 (5613, 5614)	Novel Protein sim. GBank gi 4468310 emb CAB37891  - (AL031432) dJ465N24.1 (PUTATIVE novel protein similar to predicted yeast and worm proteins) [Homo sapiens]	UNCLASSIFIED	52644507, 52645156, 52646842, 18108398, 56182575, 22278894, 22278995, 22278996, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 35696052, 29331828, 33656970, 264100, 264105, 264907, 52644045, 60433356, 264594, 60433438, 52646317, 21908754, 33109954, 33657084, 52644296, 87168474, 265011, 87168559, 265017, 265018, 265019, 18108351, 264448, 264763, 264687, 18108372, 264448, 264763, 264687, 21906768, 21906769, 35695917, 265021, 52644229, 21906765, 21906768, 21906767, 265022, 52644150, 33657023, 52645128, 33657109, 33657182, 27486261, 27486262, 27486264, 27486265, 35695763, 18108370, 18108372, 18108374, 18108376, 263977, 18108377, 35696423, 35695855, 52644332, 83373044, 18108385, 18108387, 87168518, 60432113, 60432289, 264510, 265010, 265017, 265018, 264681, 264686, 265021, 264690, 22279000, 264566, 18108351
2808	95359111 (5615, 5616)	Novel Protein sim. GBank gi 5541863 emb CAB51071.1  - (AL096857) hypothetical protein [Homo sapiens]	MHC	
2809	88083530 (5617, 5618)	Novel Protein sim. GBank gi 2772561 (AC004002) - similar to ciliary dynein beta heavy chain; 78% Similarity to P23098 (PID: g118965) [Homo sapiens]	ATPase-associated	
2810	67259032 (5619, 5620)		UNCLASSIFIED	264569, 22278996, 264091, 264259, 29331828, 29146499, 29146629, 29148784, 264693
2811	91235845 (5621, 5622)		UNCLASSIFIED	264108
2812	88093334 (5623, 5624)	Novel Protein sim. GBank gi 3264583 (AC005189) - match to ESTs H97758 (NID: g1118643) and AA085546 (NID: g1828773) [Homo sapiens]	UNCLASSIFIED	
2813	91218755 (5625, 5626)	Novel Protein sim. GBank gi 4240273 dbj BAA74915.1  - (AB020699) KIAA0892 protein [Homo sapiens]	UNCLASSIFIED	
2814	90980906 (5627, 5628)	Novel Protein sim. GBank gi 3548791 (AC005620) - R33590.1 [Homo sapiens]	transcript factor	56182575, 29331828, 264906, 66712502, 55811366, 265017, 265018, 265019, 264683, 265020, 87188518, 80432113, 65274572
2815	78774521 (5628, 5630)		UNCLASSIFIED	264907, 264909
2816	95358239 (5631, 5632)	Novel Protein sim. GBank gi 5420389 emb CAB46680.1  - (AJ243460) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	264488, 35696286, 29331825, 29331828, 264508, 264909, 264905, 264906, 264907, 264908, 264909, 264511, 264910, 265011, 264760, 264681, 264766, 264769, 264689, 21906765, 264693, 264628, 18108370, 264629, 264631, 264634, 264563, 264564, 264566, 264486



2817	87749542 (5633, 5634)	Novel Protein sim. GBank gij1293846 (J56966) - coded for by <i>C. elegans</i> cDNA yk30b3.5; coded for by <i>C. elegans</i> cDNA yk30b3.3 [Caenorhabditis elegans]			284259, 29331822, 29331827, 284503, 284508, 284905, 264907, 264908, 264909, 56182435, 264510, 265006, 264511, 264512, 264593, 264758, 265010, 264760, 264761, 264764, 264288, 264687, 264769, 55811957, 35695917, 33657109, 263978, 264634, 264636, 264639, 264584, 264585, 264566, 264488, 264587, 68712502
2818	88073578 (5635, 5636)	Novel Protein sim. GBank gij549986 (U13149) - possible apospory-associated protein [Penicillium citrare]	Contains protein domain (PF00023) - Ark repeat	UNCLASSIFIED	
2819	87793527 (5637, 5638)			UNCLASSIFIED	264809, 284511
2820	87765744 (5639, 5640)	Novel Protein sim. GBank gij4929773j9jAAD34147.1 (AF152097) CGI-05 protein [Homo sapiens]	Contains protein domain (PF00919) - Uncharacterized protein family UPF0004	UNCLASSIFIED	18108384, 52646365, 52644045, 264112, 265009, 21908754, 265017, 18108351, 264683, 264389, 264686, 264769, 21906769, 33657023, 33657109, 18108370, 18108374, 18108377, 18108385, 264638
2821	95320511 (5641, 5642)	Novel Protein sim. GBank gij398144j9jP02747C1QC_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR	Contains protein domain (PF00386) - C1q domain	complement	
2822	94260221 (5643, 5644)	Novel Protein sim. GBank gij2224671j9jBAA20820j - (AB002363) KIAA0365 [Homo sapiens]		UNCLASSIFIED	264259, 29331822, 60432289, 284905, 60433356, 33657402, 265011, 264760, 21906765, 60170615, 284838, 264766
2823	95320513 (5645, 5646)	Novel Protein sim. GBank gij398144j9jP02747C1QC_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR	Contains protein domain (PF00386) - C1q domain	complement	
2824	95320515 (5647, 5648)	Novel Protein sim. GBank gij399144j9jP02747C1QC_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR		UNCLASSIFIED	264907
2825	19742170 (5649, 5650)				264760
2826	94311905 (5651, 5652)	Novel Protein sim. GBank gij3859683jembCAA22020j - (AL033503) conserved hypothetical protein [Candida albicans]			52644156, 22278994, 35696286, 22278997, 22278998, 52645080, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 33656970, 52644045, 52646317, 33657084, 52644296, 265017, 265018, 265019, 264288, 21906764, 21906765, 21906766, 21906767, 21906769, 35695917, 52644150, 33657109, 33657182, 27486261, 27486262, 33657349, 27486265, 35695763, 35696423, 35695955, 52644332, 18108385, 87168518, 264484

2827	95320519 (5653, 5654)	Novel Protein sim. GBank gij399144jsq P02747 C10C_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR	Contains protein domain (PF00386) - complement C1q domain	264488, 263994, 264489, 65274572, 29331822, 6671417, 29331827, 29331828, 264508, 264905, 264509, 264806, 264907, 264908, 264909, 264510, 285008, 264511, 285007, 264512, 285008, 264910, 264591, 60432229, 264592, 264593, 264595, 264596, 21908754, 265011, 264600, 264601, 264602, 265017, 264604, 264605, 264288, 264766, 264767, 264689, 55811857, 264534, 264690, 264691, 264692, 264693, 263972, 264829, 35695855, 264631, 264634, 264635, 264555, 264636, 264637, 264558, 264638, 264558, 83373044, 18108385, 60432113, 22279002, 264583, 264584, 264585, 264586, 264587, 264486, 18108391
2828	91220815 (5655, 5656)	Novel Protein sim. GBank gij3598974 (AF077000) - protein tyrosine phosphatase TD14 [Rattus norvegicus]	Contains protein domain (PF00102) - phosphatase Protein-tyrosine phosphatase	29331822, 35696052, 264104, 264908, 265007, 264591, 265010, 265011, 265019, 264766, 264686, 55811857, 18108370, 18108374, 55810764, 35696423, 55811576, 58182323, 83373044, 87168518
2829	87651244 (5657, 5658)	Novel Protein sim. GBank gij4680689 gijA027734.1 AF13295 - (AF132959) C9orf25 protein [Homo sapiens]		22278996, 22278997, 264091, 264093, 60432049, 264259, 29331822, 29331825, 29331827, 29331828, 264905, 264509, 66712502, 264510, 264511, 264593, 60433438, 21908754, 265011, 264603, 18108351, 264288, 21908765, 21908768, 21908769, 29148629, 52644150, 264693, 33657109, 18108374, 264634, 18108385, 60432113, 22279000, 264585, 264488, 265008, 265019, 264639, 22279002
2830	88087109 (5659, 5660)	Novel Protein sim. GBank gij2498667 sp Q61200 NPH1_MOUSE - NEUREXOPHILIN 1		
2831	87614717 (5661, 5662)		UNCLASSIFIED	265017
2832	87631809 (5663, 5664)		UNCLASSIFIED	22278997, 22278999, 52646317, 264288, 264688, 21908767, 60431528, 264638, 22279000
2833	87612938 (5665, 5666)	Novel Protein sim. GBank gij5262615 emb CAB45747.1  - (AL080158) hypothetical protein [Homo sapiens]	UNCLASSIFIED	264555, 264556, 264558
2834	86974703 (5667, 5668)	Novel Protein sim. GBank gij2224567 db BAA20772  - (AB002311) KIAA0313 [Homo sapiens]		263972
2835	87775712 (5669, 5670)	Novel Protein sim. GBank gij4589532 db BAA76788.1  - (AB023161) KIAA0944 protein [Homo sapiens]	ATPase associated	60432289, 29331828, 265008, 265010, 265017, 264448, 55811857, 265020, 18108370
2836	85724748 (5671, 5672)	Novel Protein sim. GBank gij2351568 (U78518) - N-RAP [Mus musculus]	Contains protein domain (PF00412) - transcript factor LIM domain containing proteins	264259, 264112, 265010, 264762, 264764, 263974, 264555, 264558, 264559

2837	87765482 (5673, 5674)	Novel Protein sim. GBank gi 5420387 emb CAB46678.1  - (AJ243459) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	18108394, 22278987, 22278988, 28331822, 284907, 284909, 285006, 285007, 285018, 285019, 284682, 284288, 21906766, 21906767, 55811957, 35695917, 18108374, 56182323, 22279000, 22279002
2838	87775392 (5675, 5676)	Novel Protein sim. GBank gi 973378 (U31263) - core protein [Hepatitis C virus]	UNCLASSIFIED	18108394, 18108397, 284259, 29331826, 285007, 285019, 284448, 18108368, 20281149, 264565, 264567
2839	85799317 (5677, 5678)	Novel Protein sim. GBank gi 157515 (U64899) - thrombospondin-related anonymous protein [Plasmodium gallinaceum]	UNCLASSIFIED	284555
2840	87774665 (5678, 5680)	Novel Protein sim. GBank gi 157515 (U64899) - thrombospondin-related anonymous protein [Plasmodium gallinaceum]		284509, 284511, 285011, 284288, 284769, 285020, 284634, 284638, 284556
2841	86982568 (5681, 5682)	Novel Protein sim. GBank gi 2224603 dbj BAA20790  - (AB002330) KIAA0332 [Homo sapiens]	dna_ma_bind	56182575, 35696052, 264907, 264908, 264909, 264910, 284593, 284595, 264766, 285022, 284691, 33657182, 35695783, 18108370, 35695855, 284631, 284559, 264563, 264567
2842	80080086 (5683, 5684)	Novel Protein sim. GBank gi 5578957 emb CAB51350.1  - (AL050306) dJ47587.1 (novel KIAA0815 and KIAA0323 LIKE protein) [Homo sapiens]	UNCLASSIFIED	284600
2843	91012494 (5685, 5686)	Novel Protein sim. GBank gi 5578957 emb CAB51350.1  - (AL050306) dJ47587.1 (novel KIAA0815 and KIAA0323 LIKE protein) [Homo sapiens]		284906, 284907, 284908, 284909, 284910, 284784, 35695855, 83373044, 18108385
2844	56731154 (5687, 5688)	Novel Protein sim. GBank gi 585123 iso C08878 FBL.C. MOUSE - FIBULIN-1, ISOFORM C PRECURSOR [BASEMENT-MEMBRANE PROTEIN 90] (BM-90)	Contains protein domain (PF00009) - EGF-like domain	284685
2845	94321719 (5689, 5690)	Novel Protein sim. GBank gi 5420387 emb CAB46678.1  - (AJ243459) proteophosphoglycan [Leishmania major]	homeobox	29146498, 87168474, 264686, 35696423, 83373044, 264584
2846	88318613 (5691, 5692)	Novel Protein sim. GBank gi 5306263 gb AAD41995.1 AC006233 - (AC006233) unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	28331830, 284909, 285008, 285011, 87168559, 264629, 264556
2847	81811757 (5693, 5694)	Novel Protein sim. GBank gi 3398678 (AC005930) - R31180.1 [Homo sapiens]	UNCLASSIFIED	284908, 264766, 284769, 284629, 284637, 284566
2848	87612943 (5695, 5696)	Novel Protein sim. GBank gi 5262615 emb CAB45747.1  - (AL080156) hypothetical protein [Homo sapiens]	UNCLASSIFIED	284490, 264259, 264508, 264905, 264907, 264510, 285007, 285008, 284991, 284992, 284593, 284594, 284595, 55812038, 285010, 285011, 264604, 264763, 264784, 264785, 284788, 284888, 284628, 284629, 284555, 284636, 264556, 264557, 264638, 264558, 264559, 264563, 264566, 264567
2849	86084283 (5697, 5698)	Novel Protein sim. GBank gi 3342218 (AC004131) - Unknown gene product [Homo sapiens]		29331822, 35696052, 284509, 284906, 285007, 264594, 285018, 264288, 263972, 35696423, 18108384, 56326488, 18108390
2850	87623636 (5699, 5700)		UNCLASSIFIED	35696286, 285008, 285009, 285018, 284288, 35695917, 284693, 18108374, 35695855, 22279000

2851	87820548 (5701, 5702)	Novel Protein sim. GBank gij4321619[gb AAD15788.1] - (AF051098) seven transmembrane domain orphan receptor [Mus musculus]		UNCLASSIFIED	264906, 264907, 56182435, 264758, 55811386, 265010, 18108351, 264448, 264389, 21908785, 264691, 264692, 264693, 35695855, 264556, 18108385, 264567, 264581
2852	86987023 (5703, 5704)	Novel Protein sim. GBank gij1825729 (U88308) - similar to drosophila membrane protein PATCHED SP-P18502 (PID:g129645) [Caenorhabditis elegans]		UNCLASSIFIED	
2853	87784830 (5705, 5706)	Novel Protein sim. GBank gij2702347 (AF027503) - putative membrane-associated guanylate kinase 1 [Mus musculus]	Contains protein domain (PF00397) - WW domain		55182575, 55811150, 264680, 27486262, 27486263, 264632, 56182323, 36528486, 22279002
2854	86083557 (5707, 5708)	Novel Protein sim. GBank gij2795825 (AC004021) - kelch protein, ring canal component involved in cytoplasmic bridges: 77% Similarity to A45773 (PID:g1079096) [Homo sapiens]	Contains protein domain (PF01344) - dna_rna_bind		35696286, 29331824, 29331826, 29331828, 264908, 264768, 264693, 22279002, 264482
2855	94723856 (5709, 5710)	Novel Protein sim. GBank gij1504040[dbj BAA13219] - (D85983) similar to D.melanogaster peroxidasin(U11052) [Homo sapiens]	Contains protein domain (PF00560) - glycoprotein		22278994, 29331822, 29331824, 29331825, 264906, 264908, 265009, 33109954, 285018, 265019, 264448, 21908765, 265020, 264680, 27486265, 83373044, 22279000, 22279002, 264482
2856	88093359 (5711, 5712)	Novel Protein sim. GBank gij3264583 (AC005189) - match to ESTs H87758 (NID:g1118643) and AA085546 (NID:g1628773) [Homo sapiens]			21908766, 22278997, 265022, 29331822, 29331826, 27486282, 265007, 265009, 265017, 264482, 264583, 18108351
2857	95348288 (5713, 5714)	Novel Protein sim. GBank gij3041855 (AC004537) - similar to tumor suppressor p33ING1; similar to AF044076 (PID:g2828208) [Homo sapiens]	Contains protein domain (PF00628) - struct		22278995, 35696286, 29331824, 29331825, 35696052, 264103, 264108, 56182435, 21908765, 21908769, 265020, 18108368, 35693763, 22279002, 264583
2858	87434748 (5715, 5716)	Novel Protein sim. GBank gij462585[sp P35227 ME18_HUMAN - DNA-BINDING PROTEIN MEL-18 (ZINC FINGER PROTEIN 144)]	Contains protein domain (PF00097) - dna_rna_bind		264559, 284887, 22278995, 22278996, 22278997, 22278998, 264259, 29331826, 29331827, 29331828, 264509, 264905, 264906, 29331830, 264808, 52844045, 264909, 264511, 264512, 265007, 285008, 264910, 265009, 264593, 60433356, 264595, 264758, 21908754, 265010, 265011, 284604, 265018, 264760, 18108331, 264763, 264892, 264764, 264765, 264288, 264369, 264685, 264766, 264768, 18108337, 264769, 21908766, 21908767, 265021, 264534, 60170615, 264691, 264692, 18108370, 284629, 18108374, 264631, 264636, 263981, 18108381, 264538, 18108385, 22279002, 264554, 264566, 264486, 264567
2859	90837675 (5717, 5718)	Novel Protein sim. GBank gij4325320[gb AAD17331.1] - (AF124427) claudin-15 [Mus musculus]	UNCLASSIFIED		60424179, 65274572, 29331828, 264905, 264511, 264758, 265011, 21906767, 21908769, 55811957, 265021, 56182323

2860	87532588 (5719, 5720)	Novel Protein sim. GBank gij4469186[emb/CAB38414.1] - (AL031588) dJ1163J1.2.1 (novel protein similar to C. elegans B0035.16 and bacterial RNA (5-Methylaminomethyl-2-thiouridyate)-Methyltransferases) (isoform 1) [Homo sapiens]		UNCLASSIFIED	284256, 28331822, 29331824, 29331825, 29331826, 28331827, 29331828, 284510, 284511, 33109854, 18108351, 284683, 284765, 284369, 284686, 21908785, 284691, 284692, 284693, 18108388, 22279002, 284482
2861	86698507 (5721, 5722)	Novel Protein sim. GBank gij3941730 (AF108083) - BS4 [Homo sapiens]		UNCLASSIFIED	284368, 284692
2862	87569585 (5723, 5724)	Novel Protein sim. GBank gij4505013[ref/NP_002310.1]pLRN - leucine-rich neuronal protein		UNCLASSIFIED	284691, 284638
2863	81220421 (5725, 5726)	Novel Protein sim. GBank gij3249088 (AC004473) - Contains similarity to goliath protein gb[M87204 from D. melanogaster. (Arabidopsis thaliana)]	Contains protein domain (PF00096) - struct Zinc finger, C2H2 type		56994075, 35696286, 22278998, 29331822, 29331824, 35696052, 29331828, 284106, 284511, 55812038, 33657084, 55811386, 285018, 285019, 21906765, 21906766, 21906769, 35695917, 285020, 285022, 33657023, 33657109, 33657349, 284828, 18108376, 60431850, 56182323, 18108385, 18108387, 87168518, 22279002
2864	87420030 (5727, 5728)	Novel Protein sim. GBank gij1079451[pir/A55463 - tropomodulin, skeletal muscle - chicken]	struct		284259, 284910
2865	95312191 (5729, 5730)	Novel Protein sim. GBank gij438840 (L18048) - MSA-2 [Plasmodium falciparum]			22278895, 21906764, 284482
2866	95105480 (5731, 5732)	Novel Protein sim. GBank gij585703[sp/Q07066]PMP2_RAT - 22 KD PEROXISOMAL MEMBRANE PROTEIN	glycoprotein		65274572, 22278996, 22278998, 22278999, 284259, 29331824, 29331827, 29331828, 60433438, 21906754, 265018, 284448, 284784, 52844150, 83373044
2867	86908001 (5733, 5734)	Novel Protein sim. GBank gij4580997[gb/A024571.1]AF12108 - (AF121081) cAMP inducible 2 protein [Mus musculus]	UNCLASSIFIED		284488, 284788, 21906788, 22278998, 265023, 284259, 284508, 284605, 284907, 284511, 284910, 284635, 284636, 284637, 265011, 265017, 265018, 265018, 284583, 284088, 284566, 284764, 284368, 284567, 284485, 284288, 284768

2868	95303283 (5735, 5738)	Novel Protein sim. GBank gi 1282860 emb CAA63923  - (X94232) t-Cell activation protein [Homo sapiens]			18108392, 264488, 22278994, 22278997, 22278998, 22278999, 60432049, 284259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 35696052, 29331828, 284508, 52644045, 284828, 285006, 285007, 285008, 285009, 284591, 60432228, 284593, 60433356, 264595, 21906754, 265017, 265019, 264682, 264369, 21906765, 21906766, 21906767, 21906768, 285021, 285022, 264691, 33657182, 18108388, 27486261, 27486282, 27486284, 27486285, 18108370, 18108374, 35696423, 35695855, 284632, 56182323, 67188518, 284404, 22279000, 22279002, 264482, 284563, 284584, 264587, 264487
2869	88094412 (5737, 5738)			UNCLASSIFIED	284389
2870	84404574 (5739, 5740)			UNCLASSIFIED	284805, 264808, 264764, 21906769, 264634
2871	88318621 (5741, 5742)	Novel Protein sim. GBank gi 5306263 gb AAD41995.1 AC00623 - (AC00623) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	284259, 29331822, 60432289, 29331827, 284907, 265008, 265017, 265018, 264682, 264764, 18108354, 265021, 27486265, 284629, 18108387, 264567
2872	95312197 (5743, 5744)	Novel Protein sim. GBank gi 112205 pir B39066 - proline-rich protein 15 - rat		kinase	263981
2873	88094252 (5745, 5746)			UNCLASSIFIED	284488, 18108374, 284768, 284687, 284688, 284689, 35696423, 35696286, 35695917, 264510, 264511, 265007, 264512, 265008, 264910, 264534, 264634, 264635, 264555, 264592, 264259, 264558, 60433438, 60432289, 35696052, 265011, 264600, 264601, 60432113, 264508, 264563, 264482, 284509, 264905, 284908, 284584, 18108351, 284763, 18108370, 284807, 284586, 284808, 284764, 264288, 264587, 264809, 264486, 264766, 18108391

2874	94313548 (5747, 5748)	Novel Protein sim. GBank gij3212854 (AC004005) - unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	52644507, 52645156, 52646365, 56182575, 22278994, 22278995, 56994075, 35596286, 60432049, 264259, 52645080, 29331822, 56182181, 29331824, 60424269, 29331825, 66714117, 29331826, 29331827, 35596052, 29331828, 33656970, 66712502, 52644045, 265007, 265009, 60433356, 284758, 55812038, 18108348, 52646317, 33109954, 33657084, 265017, 264604, 265018, 285019, 284682, 264369, 264288, 264688, 52644228, 21906768, 21906767, 21906768, 55811857, 265020, 265021, 265022, 52644150, 33657023, 52645129, 18108374, 18108376, 35596423, 56182323, 18108387, 87168518, 60432113, 22279000, 22279002, 264563, 264565
2875	88083726 (5749, 5750)	Novel Protein sim. GBank gij2781366 (AC004010) - similar to Leucine-rich transmembrane proteins. 44% similarity to U42767 (P1D:g1736918) [Homo sapiens]	Contains protein domain (PF003560) - glycoprotein Leucine Rich Repeat	22278996, 22278997, 22278999, 29331828, 29331828, 29146499, 66712502, 265008, 265017, 18108351, 264683, 264689, 21906767, 18108376, 18108377, 55811576, 60170394, 22279000, 264487
2876	88080854 (5751, 5752)	Novel Protein sim. GBank gij2979530 (AC004449) - R33883_2 [Homo sapiens]	Contains protein domain (PF00167) - Igf	
2877	94747029 (5753, 5754)	Novel Protein sim. GBank gij4704208 [emb] (CA841646.1) - (AL035419) dJ1100H13.1 (putative novel protein) [Homo sapiens]	Fibroblast growth factor	52646365, 65274572, 56182575, 22278997, 22278998, 22278999, 60432049, 52645080, 60424269, 60432289, 29331827, 35596052, 29331828, 66712502, 52644045, 56182435, 60433356, 33657402, 33657084, 265019, 55811150, 264448, 264369, 21908766, 21906768, 21906769, 265020, 33657023, 33657109, 33657182, 27486262, 264629, 60431528, 55811576, 52644332, 56182323, 264558, 83373044, 18108385, 56528486, 60432113, 22279000, 22278002, 264563
2878	88095309 (5755, 5756)	Novel Protein sim. GBank gij3976775 [emb] (CA803067) - (Z81077) predicted using GeneFinder. Similarity to Yeast protein 8248 (TIR:G587531) [Caenorhabditis elegans]	UNCLASSIFIED	264488, 264259, 29331822, 29331826, 264905, 264509, 264907, 264909, 264510, 265006, 264511, 264512, 33657402, 264758, 21906754, 18108351, 264681, 264682, 264288, 264884, 264685, 264768, 264689, 21908769, 264690, 33657023, 264693, 18108364, 33657109, 264629, 18108374, 264630, 264632, 264556, 264637, 264639, 264558, 18108385, 18108387, 264563, 264564, 264565, 264566, 264488, 264567

2879	87689122 (5757, 5766)	Novel Protein sim. GBank gij4895145[gb AAD32752.1] - (AF127374) unknown [Streptomyces lavendulae]	Contains protein domain (PF00315) - UNCLASSIFIED Uracil-DNA glycosylase	18108359, 264259, 264905, 18108370, 264629, 264908, 264909, 18108374, 18108377, 265006, 264910, 264637, 60170384, 264559, 265017, 264584, 264565, 264587, 264684, 264369
2880	84851438 (5759, 5760)	Novel Protein sim. GBank gij4880703[gb AAD27741.1] (AF132966) CGI-32 protein [Homo sapiens]		264488, 52045385, 52846842, 22278994, 35898286, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35896052, 29331828, 264107, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 52644045, 264510, 265006, 264511, 265007, 264512, 265009, 264910, 264594, 21905754, 52846317, 52844286, 87168559, 284600, 284604, 284605, 264760, 264764, 264288, 264766, 264768, 264687, 264769, 21905766, 21905769, 35895917, 265021, 264690, 264692, 33857023, 52645129, 33657109, 33657182, 27486262, 33857348, 284828, 18108374, 35895885, 284634, 264635, 264636, 264637, 264638, 264557, 52644332, 264558, 264559, 83373044, 264404, 22279000, 264563, 264483, 264567, 264486
2881	87650539 (5761, 5762)	Novel Protein sim. GBank gij733571 (U23452) - No definition line found [Caenorhabditis elegans]	UNCLASSIFIED	22278998, 29331822, 52644043, 21906765, 264639, 60432113
2882	87714367 (5763, 5764)	Novel Protein sim. GBank gij1118112 (U41559) - No definition line found [Caenorhabditis elegans]		264488, 22278998, 22278999, 29331822, 29331826, 264908, 60170831, 60433356, 55812038, 264681, 264682, 264686, 264687, 264688, 21905769, 21905769, 264693, 263967, 18108374, 55811576, 56182323, 22279002, 264586
2883	95362875 (5765, 5766)	Novel Protein sim. GBank gij488008[gb AAD31087.1] (AF106934) vitamin D receptor-interacting protein [Homo sapiens]	Contains protein domain (PF00400) - UNCLASSIFIED WD domain, G-beta repeat	264112, 264682
2884	87784643 (5767, 5768)	Novel Protein sim. GBank gij2224697[dbj BAA20832] - (AB002376) KIAA0378 [Homo sapiens]	UNCLASSIFIED	265018, 264634
2885	83006308 (5769, 5770)	Novel Protein sim. GBank gij1255889 (U53344) - T07H6.5 gene product [Caenorhabditis elegans]	UNCLASSIFIED	264686, 264693
2886	91237823 (5771, 5772)	Novel Protein sim. GBank gij3882323[dbj BAA34521.1] - (AB018344) KIAA0801 protein [Homo sapiens]	Contains protein domain (PF00084) - complementrecept Sushi domain (SCR repeat)	60432049, 264259, 29331828, 264908, 264511, 264595, 60433438, 264598, 265017, 264605, 263989, 263972, 264555, 83373044, 87168518, 264586
2887	81227860 (5773, 5774)	Novel Protein sim. GBank gij3882323[dbj BAA34521.1] - (AB018344) KIAA0801 protein [Homo sapiens]	UNCLASSIFIED	18108351, 284686, 264629, 264631, 264639, 83373044, 264482
2888	95105816 (5775, 5776)	Novel Protein sim. GBank gij4508015[ref NP_003447.1] pZNF2 - zinc finger protein 205 [Homo sapiens]	Contains protein domain (PF00096) - transcriptfactor Zinc finger, C2H2 type	264488, 264259, 29331828, 264508, 264906, 264593, 284758, 264766, 264769, 18108374, 83373044, 264486



2889	87606562 (5777, 5778)	Novel Protein sim. GBank gi 85405 emb CAA58337  - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	55994075, 29331824, 265009, 264760, 18108354, 264288
2890	78703853 (5778, 5780)	Novel Protein sim. GBank gi 3877750 emb CAB01508  - (Z78064) predicted using GeneFinder; similar to collagen; cDNA EST EMBL:D65865 comes from this gene; cDNA EST EMBL:D69451 comes from this gene; cDNA EST EMBL:D68028 comes from this gene; cDNA EST EMBL:D69658 comes from this gene...		UNCLASSIFIED	264591, 264595, 264389, 264685, 264693, 264628, 264563, 264566
2892	95419745 (5783, 5784)	Novel Protein sim. GBank gi 4929759 gb A034140.1 AF15190 - (AF151903) CGI-145 protein [Homo sapiens]		UNCLASSIFIED	65274572, 35696286, 29331828, 264110, 265009, 60433438, 265018, 265019, 18108351, 264288, 21906765, 21906766, 21906769, 55811957, 264690, 65274620, 263967, 35695855
2893	87798014 (5785, 5786)			UNCLASSIFIED	284259, 265010, 264682, 18108370, 264555, 264556
2894	87755985 (5787, 5788)	Novel Protein sim. GBank gi 5689015 gb A048135.1  - (AF080171) zinc finger protein ZNF232 [Homo sapiens]	Contains protein domain (PF00086) - Zinc finger, C2H2 type	transcript factor	284259, 265008, 60433438, 52844286, 265011, 264369, 35695817, 18108381, 18108382, 18108388
2895	86938778 (5789, 5790)	Novel Protein sim. GBank gi 3924708 emb CAA84646  - (Z35397) Weak similarity with sea squirt nidogen precursor protein (blastp score 71); cDNA EST EMBL:T02069 comes from this gene; cDNA EST EMBL:D78135 comes from this gene; cDNA EST EMBL:D73147 comes from this gene; cDNA EST EMBL...	Contains protein domain (PF01437) - Plexin repeat		29331824, 265007, 264762, 264636, 264563
2896	87752122 (5791, 5792)	Novel Protein sim. GBank gi 4885548 ref NP_005456.1 pPKBG - protein kinase B gamma	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	18108392, 18108394, 18108398, 22278998, 264259, 29331822, 29331824, 29148499, 264906, 264908, 265007, 265009, 265018, 265019, 264389, 264685, 264688, 21906766, 265021, 264693, 33657182, 264839, 18108384, 18108388, 264567
2897	95413057 (5793, 5794)	Novel Protein sim. GBank gi 4502877 ref NP_001296.1 pCLDN - Clostridium perfringens enterotoxin receptor 1		UNCLASSIFIED	60424179, 58182575, 22278995, 35696286, 22278997, 22278999, 60432049, 264259, 29331824, 29331825, 60424269, 60432289, 29331826, 29331828, 35696052, 264908, 56182435, 265009, 264910, 60170831, 60431735, 60433356, 60433438, 65274444, 55811386, 265018, 18108351, 264448, 264288, 264687, 21906765, 21906766, 21906767, 21906769, 55811957, 35695917, 264634, 33657023, 33657109, 35695763, 264628, 264629, 60431528, 18108374, 55810764, 55811576, 35696423, 35695855, 264555, 56182323, 18108385, 264404, 22279000, 22279002, 264596
2898	87750340 (5795, 5796)			UNCLASSIFIED	22278995, 264604, 18108385, 264566

2899	80357670 (5787, 5788)	Novel Protein sim. GBank gij4531470[emb]CAB40137.1] - (Y18483) SLC7A8 protein [Homo sapiens]		UNCLASSIFIED	284784, 21908784, 284692
2900	94233538 (5789, 5800)			glycoprotein	65274572, 56182575, 35696286, 60432049, 284259, 29331824, 68714117, 29331828, 35698052, 29331828, 68712502, 56182435, 285008, 265007, 265008, 265009, 60433356, 284758, 265018, 284764, 284765, 284288, 284788, 21908784, 21908789, 21908789, 265020, 284692, 284693, 32833986, 284631, 83373044, 60432113
2901	87444731 (5801, 5802)	Novel Protein sim. GBank gij4759272[ref]NP_004614.1[ptTTC4 - tetrairicopeptide repeat domain 4		phosphatase	22278895, 22278897, 22278899, 60432049, 29331822, 29331824, 29331825, 29331827, 35696052, 33656970, 284910, 265009, 21908754, 33657084, 87168474, 285010, 265018, 21908784, 21908785, 21908788, 21908789, 21908789, 33657023, 284693, 33657109, 33657349, 35696423, 35695855, 283981, 56182323, 22279002
2902	85745271 (5803, 5804)	Novel Protein sim. GBank gij2414615[emb]CAB16384] - (Z99259) hypothetical protein [Schizosaccharomyces pombe]			284683, 284691
2903	87608733 (5805, 5806)	Novel Protein sim. GBank gij1079318[pir]S52241 - XLC12 protein - African clawed frog			284887, 22278894, 284259, 29331826, 29331828, 284805, 52844045, 56182435, 264511, 265017, 265018, 18108351, 284448, 264683, 284788, 284689, 35695917
2904	86458072 (5807, 5808)	Novel Protein sim. GBank gij5639823[gb]AAD45885.1[AF14367] - (AF143676) multispanning nuclear envelope membrane protein nulin [Homo sapiens]		UNCLASSIFIED	52844150, 87168518, 60432113, 22278002, 52846365, 22278899, 284259, 35696052, 265011, 265017, 284683, 284789, 35695917, 265020, 263967, 18108374, 35695855, 284637, 284952, 18108385, 18108387
2905	84449928 (5809, 5810)	Novel Protein sim. GBank gij728837[sp]P39194[ALU7 - HUMAN - III] ALU SUBFAMILY SQ WARNING ENTRY IIII		oncogene	265009, 284681, 284682
2906	95341051 (5811, 5812)	Novel Protein sim. GBank gij4689256[gb]AAD27831.1[AF12185] - (AF12185) sorting nexin 8 [Homo sapiens]	Contains protein domain (PF00787) - PX domain	UNCLASSIFIED	22278898, 35698286, 22278968, 284259, 60432289, 29331828, 29331830, 66712502, 265009, 60170831, 33109954, 284448, 264683, 284288, 284689, 21908786, 21908787, 21908788, 55811957, 35695917, 265022, 52844150, 284691, 33657023, 284692, 284693, 35695855, 60432113, 264566

2807	91211383 (5813, 5814)	Novel Protein sim. GBank gij1707078 (U80451) - contains strong similarity to a DNAJ-like domain (PS:PS00636) [Caenorhabditis elegans]	Contains protein domain (PF00226) - eph DNAJ domain	52844507, 56182575, 56181686, 22278995, 56984075, 35698286, 60432049, 58182181, 35696052, 60431735, 284595, 55812038, 21906754, 55811386, 265019, 284682, 264369, 56181582, 21906766, 55811957, 35695917, 265020, 265021, 33657023, 33657109, 60431528, 55811576, 35696423, 35695855, 264638, 22279000
2808	80414246 (5815, 5816)	Novel Protein sim. GBank gij2673917 (AC002561) - putative ATP-dependent RNA helicase [Arabidopsis thaliana]	helicase	265009, 33109954, 18108351, 264766, 265021, 284691, 284692, 18108374, 284556, 284638, 284557, 284558
2809	87420225 (5817, 5818)		eph	264259, 87168474, 265018, 18108365, 284628
2910	86601075 (5819, 5820)	Novel Protein sim. GBank gij4539335[emb]CAB37483.1] - (AL035539) putative protein [Arabidopsis thaliana]		22278995, 264509, 264512, 265007, 33657402, 265017, 264369, 265022, 18108385, 284628
2911	94216615 (5821, 5822)	Novel Protein sim. GBank gij4469187[emb]CAB38415.1] - (AL031588) dJ1163J1.3 (novel protein similar to mouse B99) [Homo sapiens]	glucoamylase	52646365, 18108397, 22278995, 22278997, 22278998, 22278999, 29331824, 29331825, 52644045, 265006, 265018, 264448, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265021, 18108370, 18108372, 18108374, 22279000
2912	87731803 (5823, 5824)	Novel Protein sim. GBank gij4929637[gb]AAD34078.1[AF15184] - (AF151842) CGI-84 protein [Homo sapiens]	Contains protein domain (PF00904) - Involucrin repeat	52645156, 264082, 60432049, 264259, 52645080, 29331824, 29331825, 66712502, 33109954, 264760, 264883, 264288, 284686, 265021, 284693, 18108398, 263976, 264404
2913	87713823 (5825, 5826)	Novel Protein sim. GBank gij854065[emb]CAA58337] - (X83413) U88 [human herpesvirus 6]	UNCLASSIFIED Zinc finger, C2H2 type	52644507, 52645156, 52646842, 56182375, 35696286, 22278997, 264259, 52645080, 29331827, 35696052, 29331828, 264828, 52644045, 56182435, 55812038, 52846317, 21906754, 52644296, 87168474, 265017, 265018, 285018, 18108351, 284682, 284686, 264689, 21906765, 21906766, 21906767, 21906768, 35695917, 265020, 52844150, 27486261, 27486262, 27486265, 35695763, 55811576, 35695855, 52644332, 22279000, 22279002, 284563
2914	87797300 (5827, 5828)			264557

2915	88081972 (5829, 5830)	Novel Protein sim. GBank gjl5174485(re)[NP_006030.1]pK1AA - endocytic receptor (macrophage mannose receptor family)	Contains protein domain (PF00059) - eph Lectin C-type domain	264569, 264488, 264687, 264768, 21806766, 52646842, 21906767, 21908768, 58182575, 29148629, 35899917, 22278996, 22278997, 22278998, 285021, 22278998, 52644150, 264691, 264259, 60432049, 264892, 52645129, 33657109, 33657182, 29331827, 27486261, 35696052, 29331828, 27486262, 27486264, 27486265, 33657349, 29146498, 29146499, 264906, 264907, 18108370, 264908, 18108372, 52644045, 18108374, 56182435, 35695855, 264112, 264510, 265008, 60432229, 264593, 60433356, 56182323, 18108382, 55812038, 18108385, 33109954, 21906754, 33657084, 87168518, 265010, 265011, 60432113, 265017, 265018, 22279000, 265019, 55811150, 264681, 18108351, 264763, 264448, 264683, 264566, 18108354, 264369, 264288, 264766, 52645156, 85274572, 22278994, 22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 66714117, 60432289, 29331826, 29331827, 29331828, 33658970, 264509, 264908, 29331830, 52644045, 264909, 56182435, 60170831, 264592, 264593, 33657402, 60433356, 52646317, 21906754, 33109954, 33657084, 52644296, 85658542, 265011, 265017, 265018, 265019, 18108351, 264448, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35899917, 265021, 265022, 52644150, 33657023, 33657109, 33657182, 27486261, 27486262, 27486284, 35695763, 18108376, 55811576, 35698423, 65274791, 35695855, 52644332, 264557, 264638, 56182323, 18108387, 87168518, 22279002, 264482
2916	95337790 (5831, 5832)	Novel Protein sim. GBank gjl5104851[db][BAA0165.1] - (AP000061) 305aa long hypothetical dTDP-4- dehydrohamose reductase [Aeropyrum pernix]	dehydrogenase	
2917	87454546 (5833, 5834)	Novel Protein sim. GBank gjl3169065[emb][CAA19260.1] - (AL023704) putative translocation elongation factor-Tu fa mily [Schizosaccharomyces pombe]	UNCLASSIFIED	60433438, 264602, 264682, 87168518, 60432113
2918	85690529 (5835, 5836)	Novel Protein sim. GBank gjl539218[pr][S38038 - hypothetical protein YKL201c - yeast (Saccharomyces cerevisiae)]		264638
2919	87641497 (5837, 5838)	Novel Protein sim. GBank gjl2584655 (AF030001) - unknown [Mus musculus]		66714117, 66712502, 263981

2920	87769523 (5639, 5840)				35695286, 22278997, 264259, 52645080, 29331824, 29331826, 29331827, 264828, 264909, 56182435, 264511, 284758, 33109954, 21908754, 52644296, 265010, 285011, 264601, 285017, 265019, 264681, 264687, 21908767, 265021, 52644150, 264680, 264691, 264692, 264693, 33657109, 33657182, 27486282, 27486284, 27486285, 35696423, 35695855, 264632, 264636, 264637, 264638, 56182323, 60170394, 18108385, 87168518, 60432113, 35695286, 22278997, 264091, 284092, 264094, 284259, 29331822, 29331824, 29331826, 29331827, 35696052, 29146498, 2641104, 264105, 264107, 264509, 264110, 264112, 264512, 60433356, 21906754, 87168474, 265017, 18108351, 264288, 21906765, 21906768, 21906769, 35695917, 265021, 263974, 18108374, 263976, 263977, 18108378, 264555, 263681, 55526466, 87168518, 22279000, 22279002, 264906, 264909, 264511, 265006, 265008, 264593, 33657402, 60174639, 18108351, 264763, 21908765, 29148627, 35695917, 264682, 264628, 263978, 55811576, 35695855, 264555, 264558, 56182323, 60170394, 22279000, 284486
2921	91639982 (5841, 5842)	Novel Protein sim. GBank gij4580013jgIAAD24202.1[U83194, associated factor 2 [Homo sapiens]	Contains protein domain (PF00787) - PX domain		264488, 18108397, 22278995, 22278996, 22278997, 22278998, 22278999, 29331825, 29331826, 29331827, 29331830, 264511, 265009, 33657402, 265011, 265017, 265018, 284683, 18108354, 21908765, 21908767, 21908768, 21908769, 52644150, 264691, 284682, 33657109, 263974, 18108376, 264631, 264636, 18108385, 18108387, 22279000, 264563, 264568, 265017, 264628, 20281152, 264556
2922	87749762 (5843, 5844)	Novel Protein sim. GBank gij4589514[dbj]BAA76775.11 - (A8023152) KIAA0935 protein [Homo sapiens]	Contains protein domain (PF01074) - Kinase Glycosyl hydrolases family 38		264488, 18108397, 22278995, 22278996, 22278997, 22278998, 22278999, 29331825, 29331826, 29331827, 29331830, 264511, 265009, 33657402, 265011, 265017, 265018, 284683, 18108354, 21908765, 21908767, 21908768, 21908769, 52644150, 264691, 284682, 33657109, 263974, 18108376, 264631, 264636, 18108385, 18108387, 22279000, 264563, 264568, 265017, 264628, 20281152, 264556
2923	95337799 (5845, 5846)	Novel Protein sim. GBank gij4835268[emb]CAB42898.21 - (Z83844) dJ37E16.4 (similar to mouse p116Rip protein) [Homo sapiens]	Contains protein domain (PF00169) - PH domain		264488, 18108397, 22278995, 22278996, 22278997, 22278998, 22278999, 29331825, 29331826, 29331827, 29331830, 264511, 265009, 33657402, 265011, 265017, 265018, 284683, 18108354, 21908765, 21908767, 21908768, 21908769, 52644150, 264691, 284682, 33657109, 263974, 18108376, 264631, 264636, 18108385, 18108387, 22279000, 264563, 264568, 265017, 264628, 20281152, 264556
2924	87791987 (5847, 5848)	Novel Protein sim. GBank gij2133095[pir]S72254 - ribosomal protein L36, mitochondrial - yeast [Saccharomyces cerevisiae]	Contains protein domain (PF00444) - Ribosomal protein L36	ribosomalprot	264488, 18108397, 22278995, 22278996, 22278997, 22278998, 22278999, 29331825, 29331826, 29331827, 29331830, 264511, 265009, 33657402, 265011, 265017, 265018, 284683, 18108354, 21908765, 21908767, 21908768, 21908769, 52644150, 264691, 284682, 33657109, 263974, 18108376, 264631, 264636, 18108385, 18108387, 22279000, 264563, 264568, 265017, 264628, 20281152, 264556
2925	95090120 (5849, 5850)	Novel Protein sim. GBank gij2388986[emb]CAB117181 - (Z98980) actin associated protein [Schizosaccharomyces pombe]	UNCLASSIFIED		56182575, 35686286, 264259, 60432289, 29331827, 264508, 52644045, 284810, 264591, 60432228, 55812038, 21906754, 284681, 264448, 264683, 264288, 264085, 52644228, 264689, 21906765, 21906766, 21906768, 21906769, 265021, 265022, 60170615, 264692, 33657023, 264693, 33657109, 35696423, 65274761, 56182323

2928	85343003 (5851, 5852)				28331828, 265011, 264768, 264689
2927	80408018 (5853, 5854)	Novel Protein sim. GBank gij283032[pir][S22455 - hydroxyproline-rich glycoprotein - perennial leosinte]			284764, 284288, 264630, 264637
2928	20452179 (5855, 5856)			UNCLASSIFIED	264559
2929	91622820 (5857, 5858)	Novel Protein sim. GBank gij3413320[emb][CA06915] - (AJ008215) CMP-N-acetylneuraminic acid synthetase [Mus musculus]		UNCLASSIFIED	264559, 284489, 22278994, 35698286, 22278998, 22278999, 264259, 29331825, 264259, 52645080, 28331822, 29331824, 66714117, 29331825, 29331826, 29331827, 35696052, 33696370, 264109, 29331830, 52644045, 265009, 33109954, 52644296, 87168559, 264760, 264762, 264448, 264764, 264288, 284766, 284768, 21908765, 21908768, 21908789, 35695917, 264691, 33657023, 264693, 33657109, 18108374, 263978, 35696423, 35695855, 263981, 22279000, 22279002, 264567, 264488
2930	95302755 (5859, 5860)			UNCLASSIFIED	56182575, 56181686, 35696286, 22278998, 22278999, 264259, 29331825, 60432289, 29331828, 264905, 52644045, 56182435, 265009, 60170831, 264592, 60432229, 60433356, 87168474, 265010, 265011, 265017, 265018, 265019, 264762, 264448, 264693, 264288, 264766, 21908765, 21908789, 35695917, 60170815, 33657023, 33657109, 284628, 18108370, 18108372, 35696423, 35695855, 264556, 56182323, 60432113, 264567
2931	94312693 (5861, 5862)	Novel Protein sim. GBank gij3786433 (AF098505) - similar to Arabidopsis thaliana male sterility protein 2 (SW:Q08891) [Caenorhabditis elegans]	Contains protein domain (PF00471) - Ribosomal protein L33	UNCLASSIFIED	52645156, 22278997, 22278998, 29331822, 52645080, 29331824, 60432289, 35696970, 60433356, 60433438, 33109954, 21908765, 21908766, 21908787, 21908789, 265020, 52644150, 33657023, 33657109, 33657182, 27486285, 35696423, 35695855, 264555, 87168518, 60432113, 264566
2932	79632623 (5863, 5864)				264906, 264907
2933	91720776 (5865, 5866)	Novel Protein sim. GBank gij3378056 (AF017777) - helicase [Drosophila melanogaster]	helicase		264488, 18108392, 56182575, 22278999, 264091, 284259, 29331825, 60432289, 29331827, 264508, 52644045, 56182435, 265007, 265009, 264592, 60433356, 60433438, 21908754, 265017, 264682, 264288, 52644229, 21908765, 21908768, 21908789, 265022, 52644150, 33657023, 33657109, 27486285, 264635, 264636, 60170394, 56182323, 18108385, 60432113, 264565, 264566, 264567

2934	86576025 (5667, 5868)				22278987, 22278999, 29331824, 33657402, 264691, 27468262, 264628, 87168518, 22279000
2935	86410579 (5869, 5870)			UNCLASSIFIED	56182575, 22278995, 60433356, 33657402, 264758, 33109954, 21908754, 265018, 265019, 264448, 264769, 21906764, 21906765, 285021, 264692, 33657023, 33657109, 33657349, 55810764, 22279000
2936	87603863 (5871, 5872)	Novel Protein sim. GBank gi14153862 (AC005065) - determined by GENSCAN prediction and spliced EST; match to EST RM4329 (NID:942735) [Homo sapiens]	Contains protein domain (PF00856) - SET domain	nuclease	22278997, 29331827, 29331828, 265009, 265017, 264605, 265020, 55811578, 18108387, 60432113, 264563
2937	94853096 (5873, 5874)	Novel Protein sim. GBank gi5174409[refNP_006101.1]pCD2B - CD2 antigen (cytoplasmic tail)-binding protein 2		UNCLASSIFIED	55994075, 22278999, 264259, 60432049, 29331822, 56182181, 29331827, 29331828, 264906, 264908, 264909, 56182435, 265006, 264512, 264910, 60170831, 60433356, 265011, 265018, 18108351, 264448, 264288, 264768, 52844229, 21906765, 29148784, 65274791, 264556, 56182323, 60170394, 264558, 60432113, 284565, 284486, 284597, 264488, 56182575, 22278996, 33696286
2938	95418773 (5875, 5876)	Novel Protein sim. GBank gi3319990[emb]CAA76720 - (Y17267) ubiquitin-conjugating enzyme [Mus musculus]	Contains protein domain (PF00179) - Ubiquitin-conjugating enzyme	ubiquitin	22278997, 22278998, 22278999, 264490, 264259, 29331822, 29331824, 66714117, 29331827, 33696052, 264107, 264905, 66712502, 52644045, 56182435, 264511, 265008, 265009, 60432228, 33657402, 60433438, 55812038, 21906754, 85658542, 265010, 265011, 87168359, 265017, 265018, 265019, 264681, 264288, 264689, 21906765, 21906767, 21906768, 55811957, 33695917, 265020, 60170815, 264690, 264691, 264692, 33657023, 264693, 65274620, 33657109, 18108370, 18108374, 263976, 33696423, 35695855, 264555, 264556, 18108381, 56182323, 60170394, 83373044, 18108385, 55526488, 60432113, 22279002
2939	97786622 (5877, 5878)	Novel Protein sim. GBank gi3379990[emb]CAA99909 - (Z75547) similar to WD domain, G-beta repeat; cDNA EST yk371b7.5 comes from this gene; cDNA EST yk312h1.5 comes from this gene; cDNA EST yk485d5.5 comes from this gene; cDNA EST yk472c4.5 comes from this gene; cDNA EST yk292b.....	Contains protein domain (PF00400) - WD domain, G-beta repeat	ATPase-associated	264907, 265018, 264681, 264685, 264686

2940	95011103 (5879, 5880)				UNCLASSIFIED	22278986, 29331822, 29331824, 66714117, 29331826, 29331828, 284905, 284908, 66712502, 29331830, 265011, 265017, 264764, 264369, 21906766, 21906767, 33657023, 33657109, 32833986, 18108374, 18108377, 264634, 83373044, 18108385, 18108387, 264566 284557
2941	21423370 (5881, 5882)	Novel Protein sim. GBank gi 3413872 dbj BAA33200  - (AB007824) KIAA0455 protein [Homo sapiens]			UNCLASSIFIED	264910, 265010, 264768
2942	87430203 (5883, 5884)	Novel Protein sim. GBank gi 1172845 sp P46629 RB25_RABIT - RAS-RELATED PROTEIN RAB-25			glycoprotein	
2943	95314504 (5885, 5886)	Novel Protein sim. GBank gi 4929653 dbj AAD34087.1 AF15185 - (AF151850) CGI-92 protein [Homo sapiens]			collagen	60432049, 264259, 60432289, 29331827, 29146496, 265008, 284593, 60433356, 60433438, 265010, 265011, 265017, 265018, 264683, 264766, 18108381, 65274727, 60432113, 264567
2944	95081063 (5887, 5888)	Novel Protein sim. GBank gi 4678282 emb CAB41180.1  - (AL049650) 1-acylglycerol-3-phosphate acyltransferase-like protein [Arabidopsis thaliana]	Contains protein domain (PF00415) - Regulator of chromosome condensation (RCC1)		ATPase-associated	56994075, 22278998, 60432049, 264259, 29331822, 29331824, 60424269, 60432289, 29331826, 29331828, 284905, 264907, 52644045, 264909, 264511, 265006, 265009, 264594, 21906754, 87188559, 264603, 265017, 265018, 18108351, 264682, 264766, 284687, 264688, 21906765, 21906766, 21906767, 21906768, 21906769, 265021, 60170615, 52644150, 284690, 264691, 33657023, 264692, 264693, 33657109, 33657182, 33657349, 18108370, 18108374, 18108377, 55811576, 35696423, 35696855, 264635, 264555, 264556, 68182323, 60170394, 264558, 264559, 83373044, 58525486, 87168518, 60432113, 22279002, 284482, 264563, 264484, 264567
2945	94233560 (5889, 5890)	Novel Protein sim. GBank gi 728831 sp P39188 ALU1_HUMAN - IIII ALU SUBFAMILY J WARNING ENTRY IIII	Contains protein domain (PF00096) - Zinc finger, C2H2 type		UNCLASSIFIED	60424179, 22278995, 22278996, 22278998, 22278999, 264259, 56182181, 29331824, 60424269, 60432289, 35696052, 284908, 265006, 60433386, 55812038, 264759, 55811386, 265018, 264681, 18108351, 284448, 264683, 264369, 264288, 264687, 55811562, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 264693, 60431528, 55810764, 35696423, 35696855, 264630, 60170394, 83373044, 22279000, 264566, 264567



2946	94317315 (5891, 5892)	Novel Protein sim. GBank gi 5441952 gb AAD43185.1 AF07288 - (AF072884) peroxisomal membrane protein PMP 24 [Homo sapiens]		UNCLASSIFIED	264488, 264259, 264508, 264509, 264806, 264907, 264809, 264510, 264511, 265007, 264612, 264810, 264891, 264593, 18108351, 264764, 264288, 264884, 264769, 265021, 264692, 33657109, 264628, 264629, 18108374, 264631, 264634, 264636, 264637, 18108380, 264638, 264639, 83373044, 264565, 264566, 264486, 264567
2947	87362952 (5893, 5894)	Novel Protein sim. GBank gi 3540281 gb AAC34383.1  - (AF056116) Ali-1 related protein [Fugu rubripes]		UNCLASSIFIED	22278955, 22278956, 22278957, 22278958, 29146498, 264508, 29331830, 265007, 265008, 265009, 60432228, 21906754, 265010, 265017, 265019, 264766, 264685, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 264628, 18108370, 264629, 264630, 18108387, 60432113
2948	87626527 (5895, 5896)	Novel Protein sim. GBank gi 5566614 gb AAB56554.2  - (AF001533) mitogen-induced [Mus musculus]			52646842, 22278955, 264259, 29331824, 28331825, 28331827, 29331830, 264809, 265007, 265009, 265018, 264763, 264684, 264288, 264685, 264686, 21906767, 264691, 264692, 264693, 18108374, 55811576, 18108385, 22279002, 264563, 264567
2949	88175545 (5897, 5898)	Novel Protein sim. GBank gi 2132823 pir S67133 - probable membrane protein YOR240w - yeast (Saccharomyces cerevisiae)		UNCLASSIFIED	22278956, 22278957, 60432269, 29331826, 28331827, 29331828, 35698052, 28148499, 264104, 264107, 264905, 60712502, 264908, 60433356, 60433438, 87168559, 264764, 52644229, 56181562, 21906767, 21906768, 21906769, 265022, 60170615, 33657023, 35696423, 263981, 264558, 60432113, 22279002
2950	95086870 (5899, 5900)	Novel Protein sim. GBank gi 466102 sp P34629 YQJ6_CAEEL - PUTATIVE AMINOPEPTIDASE ZK353.6 IN CHROMOSOME III	Contains protein domain (PF00883) - peptidase Cytosol aminopeptidase family		264488, 35696286, 264259, 35698052, 264907, 265007, 264810, 265017, 265018, 264288, 264768, 35695917, 265020, 18108362, 18108370, 18108379, 35696423, 65274791, 35695855, 264556, 56526488, 264486
2951	87392357 (5901, 5902)	Novel Protein sim. GBank gi 4688902 emb CAB41450.1  - (AJ238248) centaurin beta2 [Homo sapiens]			264893

2952	95329952 (5903, 5904)	Novel Protein sim. GBank gll5596893[emb CAE51405.1] - (AL096881) hypothetical protein [Homo sapiens]	Contains protein domain (PF00650) - CRAL/TRIO domain.	transcript factor	264687, 52645156, 21906766, 21906769, 22278998, 265020, 264690, 60432049, 264259, 264693, 29331822, 18108365, 29331825, 60432289, 33657109, 18108368, 29331827, 35696052, 27486262, 264508, 284905, 20281149, 264806, 264807, 29331830, 284908, 264809, 35695855, 264511, 265008, 265009, 264910, 264635, 284636, 60432228, 264638, 60433358, 284639, 264758, 87168518, 265017, 22279000, 22279002, 264760, 264563, 264482, 18108351, 264448, 264288
2953	80933575 (5905, 5906)	Novel Protein sim. GBank gll119522[sp P10658]SERC, RABIT - PROBABLE PHOSPHOSERINE AMINOTRANSFERASE (PSAT) (ENDOMETRIAL PROGESTERONE-INDUCED PROTEIN) (EPIP)	Contains protein domain (PF00268) - Aminotransferases class-V	UNCLASSIFIED	18108366, 55994075, 22278996, 29331822, 29331824, 29331825, 29331827, 35696052, 264508, 264905, 264908, 264907, 264510, 264591, 264594, 33657402, 264595, 264596, 264758, 52646317, 21906754, 33657084, 52644296, 87168559, 264600, 264760, 264681, 18108351, 264784, 264359, 264288, 264687, 21908785, 21908786, 21906787, 21908788, 21908789, 35695917, 33657023, 18108384, 52645129, 33657109, 33657348, 18108374, 263978, 35695855, 264637, 264638, 87168518, 264482, 264563, 264565, 29331822
2954	88086288 (5907, 5908)	Novel Protein sim. GBank gll4885261[ref NP_005251.1]pGDF9 - growth differentiation factor 9	Contains protein domain (PF00019) - Transforming growth factor beta like domain	Igf	
2955	87698426 (5909, 5910)	Novel Protein sim. GBank gll3452473 (AF084205) - serine/threonine protein kinase TAO1 [Rattus norvegicus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264259, 29331822, 29331824, 29331825, 29331826, 35696052, 264608, 52644045, 264512, 60432228, 265018, 265019, 55811150, 264769, 21906787, 21908788, 21906769, 285021, 50170615, 55810764, 264567
2956	85789745 (5911, 5912)	Novel Protein sim. GBank gll4889254[gb AA027830.1]AF12185 - (AF121857) sorting nexin 7 [Homo sapiens]	Contains protein domain (PF00787) - PX domain		22278996, 264259, 29331827, 264908, 21906768
2957	90933301 (5913, 5914)	Novel Protein sim. GBank gll4503023[ref NP_000089.1]pCPT2 - carnitine palmitoyltransferase II precursor		cadherin	22278999, 264259, 29331824, 29331827, 265008, 264595, 264758, 265010, 265011, 264448, 264763, 264683, 264288, 264685, 18108357, 29148629, 264690, 18108362, 264693, 18108370, 80431528, 18108374, 264634, 18108381, 56182323, 18108382, 18108385, 18108388, 56526486, 87168518, 264482, 264487
2958	87440014 (5915, 5916)	Novel Protein sim. GBank gll4240257[db BAA74807.1] - (AB020691) KIAA0884 protein [Homo sapiens]			264595, 264596, 264681, 264369, 264629, 264631, 264567

2959	95109420 (5917, 5918)	Novel Protein sim. GBank gij398221 (U33005) - Tbc1 [Mus musculus]	Contains protein domain (PF00566) - oncogene TBC domain	263994, 22278997, 284259, 60432049, 29331826, 29331828, 35998052, 29331830, 68712502, 56182435, 265006, 284512, 265008, 265009, 60433356, 60433438, 264596, 285017, 285018, 264683, 264288, 264768, 264769, 21906766, 21906767, 21908769, 265020, 60170615, 264692, 27486285, 18108374, 65274781, 35695855, 83373044, 56528489, 80432113
2960	87420091 (5919, 5920)		UNCLASSIFIED	35696286, 56182435, 87168474, 265010, 60170615, 35698423, 56182323, 18108383, 87168518, 264483
2961	95413416 (5921, 5922)	Novel Protein sim. GBank gij596646 (emb CA805177.2  - (Z62266) predicted using GeneFinder: similar to WD domain, G-beta repeats [Caenorhabditis elegans]	Contains protein domain (PF00400) - WD domain, G-beta repeat	22278997, 22278999, 264259, 29331822, 29331824, 29331826, 29331828, 264907, 264908, 52844045, 285006, 33657402, 21908754, 87168474, 285011, 87168559, 265017, 21906769, 265020, 60170615, 264692, 33657023, 35695783, 18108370, 18108374, 35698423, 264632, 264636, 18108385, 87168518, 22279002, 264564, 284567
2962	87912700 (5923, 5924)		UNCLASSIFIED	35696286, 22278997, 264092, 264094, 264259, 29331824, 66714117, 29331825, 60432289, 29331826, 29331827, 29331828, 35998052, 284508, 284905, 284508, 264907, 264908, 264909, 264510, 264512, 264593, 264594, 60433438, 264758, 52846317, 264602, 264603, 264605, 264760, 264762, 264764, 264288, 264766, 264686, 264768, 264769, 35695917, 265020, 264691, 264634, 264636, 264637, 264638, 264639, 18108385, 264563, 264565, 264566, 264567, 264486
2963	95313464 (5925, 5926)	Novel Protein sim. GBank gij5240223 (dbj BAA74890.1  - (AB020674) KIAA0867 protein [Homo sapiens])	Contains protein domain (PF00010) - Helix-loop-helix DNA-binding domain	18108382, 56994073, 22278998, 22278999, 29331822, 29331825, 29331826, 29331827, 29331828, 285007, 265008, 284592, 264594, 21906754, 285018, 264760, 264687, 29148627, 29148784, 285020, 33657023, 284693, 65274620, 33657182, 27486281, 264629, 55810764, 35696423, 264555, 264636, 264637, 264557, 264558, 264563
2964	94324517 (5927, 5928)		UNCLASSIFIED	264259, 29331828, 33657402, 265017, 265018, 264692, 18108388, 35698423, 83373044, 18108388

2965	80384762 (5929, 5930)	Novel Protein sim. GBank gi 4895447 ref NP_005432.1 pKRM1 - Kreisler (mouse) mal- related leucine zipper homolog			transcript factor	284259, 29331826, 264508, 284508, 284508, 284905, 264907, 264908, 264909, 264511, 265008, 264910, 264591, 264593, 264594, 33657402, 285011, 264760, 264762, 264764, 264288, 264685, 284788, 284892, 33657109, 264628, 264629, 35895855, 264630, 284631, 264632, 264634, 264635, 264636, 284637, 264638, 264639, 264583, 264567, 18108391 60432289, 264682, 264448
2966	91725248 (5931, 5932)	Novel Protein sim. GBank gi 5262751 emb CA54890.1  - (AJ243177) Xenopus RPA interacting protein alpha [Xenopus laevis]				
2967	94658303 (5933, 5934)	Novel Protein sim. GBank gi 624225 (U19181) - Rabin3 [Rattus norvegicus]			UNCLASSIFIED	264488, 264508, 284508, 264908, 264909, 284511, 264910, 264594, 264758, 85658542, 264762, 264764, 285021, 264556, 18108381, 264584, 264485
2968	95302776 (5935, 5936)	Novel Protein sim. GBank gi 4929715 gb AAD34118.1 AF15188 - (AF15188) CGI-123 protein [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C2HC4 type (RING finger)			264687, 52645156, 21908765, 52845365, 21908767, 18108398, 35896423, 22278996, 35896286, 22278997, 265020, 22278999, 265021, 265022, 264093, 264636, 264690, 52844150, 264258, 33657023, 52645080, 264693, 29331822, 56182181, 29331824, 56714117, 29331825, 33109954, 52845129, 29331826, 21908754, 33657182, 29331827, 29331828, 33698032, 27486262, 87168518, 87168474, 265010, 87168559, 265018, 22279000, 265018, 22279002, 264583, 18108351, 264906, 264907, 264448, 66712502, 264566, 264369, 264288
2969	95310957 (5937, 5938)	Novel Protein sim. GBank gi 3024743 pp O24734 THSA_SULS7 - THERMOSOME: ALPHA SUBUNIT (CHAPERONIN ALPHA SUBUNIT)			eph	52845842, 22278996, 22278998, 22278999, 60432049, 264259, 29331824, 29331825, 29331826, 29331828, 264509, 264909, 52644045, 56182435, 265009, 60433438, 55812038, 21908754, 265011, 87168559, 265018, 265018, 264448, 264288, 264369, 52644229, 21908766, 21908768, 21908769, 29148784, 265020, 265021, 52844150, 264681, 33657109, 18108374, 56182323, 60170394, 87168518, 60432113, 22279000
2970	88088071 (5939, 5940)	Novel Protein sim. GBank gi 3165407 (AC004755) - Ios37502.1 [Homo sapiens]	Contains protein domain (PF00046) - Homeobox domain		homeobox	

2971	94186930 (5941, 5942)	Novel Protein sim. GBank gij728837spjP39194ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		lm7	264488, 56182575, 35696286, 56994075, 29331824, 29331826, 29146499, 264508, 264905, 264907, 264112, 264910, 21908754, 87188559, 265018, 285019, 18108351, 264689, 21906765, 21906767, 21906768, 265020, 265021, 60170615, 18108384, 264628, 264629, 18108374, 264636, 264556, 264558, 83373044, 18108384, 18108385, 87188518, 264584, 264587, 265017, 35695917, 265021, 33657109, 22279002, 264583
2972	86325943 (5943, 5944)	Novel Protein sim. GBank gij728836spjP39193ALU6_HUMAN - IIII ALU SUBFAMILY SP WARNING ENTRY IIII		kinase	29331822, 264692, 33657349, 55811576, 264583
2973	91215301 (5945, 5946)	Novel Protein sim. GBank gij2746789 (AF040842) - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	18108392, 52644507, 56182575, 56181686, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 52645080, 29331824, 29331825, 66714117, 60424269, 29331826, 29331827, 29331828, 35698052, 66712502, 264908, 52644045, 265007, 264910, 265009, 60433438, 33109954, 21908754, 55811386, 52644286, 87168474, 87188559, 285017, 265018, 265019, 18108351, 264448, 264369, 264288, 52644229, 18108398, 21906765, 21906767, 21906768, 35695917, 265020, 265021, 52644150, 264691, 264692, 33657023, 27486282, 27486284, 35695783, 18108370, 18108376, 55810764, 55811576, 35696423, 35695855, 264630, 264635, 264557, 52644332, 284558, 83373044, 18108387, 87188518, 60432113, 22279000, 264482, 264487
2975	95325213 (5949, 5950)	Novel Protein sim. GBank gij3880812lmb(CAA19508) - (AL023839) similar to HECT-domain (ubiquitin-transferase) ; cDNA EST YK480d10.5 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00632) - HECT-domain (ubiquitin- transferase).	ubiquitin	29331824, 29331827, 29331828, 264910, 85658542, 285011, 285018, 264448, 264288, 264769, 21908767, 265020, 264681, 264559, 83373044
2976	87771202 (5951, 5952)	Novel Protein sim. GBank gij5679136gijA046874.1JAF16093 - (AF160934) BcDNA LD14189 [Drosophila melanogaster]		transport	22278996, 264906, 265007, 265010, 265011, 265017, 265018, 18108351, 264685, 264689, 18108370, 264639, 18108385
2977	91725254 (5953, 5954)	Novel Protein sim. GBank gij5262751lmb(CAB45690.1) - (AJ243177) Xenopus RPA Interacting protein alpha [Xenopus laevis]		UNCLASSIFIED	264509, 264288

2978	87332059 (5955, 5956)	Novel Protein sim. GBank gij746549 (U23522) - No definition line found [Caenorhabditis elegans]	Contains protein domain (PF00480) - ROK family	UNCLASSIFIED	22278995, 22278996, 22278987, 22278989, 264259, 60432289, 29331827, 29146489, 56182435, 265006, 265007, 265009, 60433356, 60433438, 21908754, 265010, 265011, 265017, 265018, 265019, 264288, 264685, 264688, 21908765, 21908766, 21908767, 21908769, 265020, 265021, 265022, 35686423, 264639, 60432113, 22279000, 22279002
2979	91752558 (5957, 5958)	Novel Protein sim. GBank gij5262751jemb[CAB45680.1] - (AJ243177) Xenopus RPA interacting protein alpha [Xenopus laevis]	complement		264488, 65274572, 56994075, 22278999, 264093, 29331822, 29331824, 264288, 55811957, 33657023, 33657109, 18108370, 55811576, 56182323, 60432113, 264482, 265009, 21908767, 263981, 22279000
2980	86286600 (5959, 5960)			UNCLASSIFIED	264629, 264564
2981	87376330 (5961, 5962)				
2982	95303675 (5963, 5964)	Novel Protein sim. GBank gij4929787jgbAAD34144.1[AF15190 - (AF151907) CGI-149 protein [Homo sapiens]			22278995, 56994075, 22278996, 22278997, 22278998, 22278999, 264092, 29331824, 29331827, 29331828, 264905, 264591, 264592, 264594, 264595, 264596, 33657084, 264448, 21908765, 21908766, 21908767, 21908768, 21908769, 265020, 265022, 18108365, 33657182, 33657349, 35686423, 83373044, 22279000, 22279002
2983	91752558 (5965, 5966)	Novel Protein sim. GBank gij5262751jemb[CAB45680.1] - (AJ243177) Xenopus RPA interacting protein alpha [Xenopus laevis]			60424179, 52646842, 18108398, 22278997, 264093, 60432049, 264259, 29331822, 60432289, 33656970, 264905, 52644045, 265008, 60431735, 87168474, 285018, 265019, 18108351, 264448, 21908765, 21908766, 35685917, 33657023, 52645129, 18108370, 35686423, 83373044, 56528486, 60432113, 264404, 22279002
2984	94136467 (5967, 5968)	Novel Protein sim. GBank gij2393734 (AC002542) - similar to C. elegans F1A10.5; 80% similarity to Z68297 (PID:g1130619) [Homo sapiens]	ATPase associated		
2985	87099072 (5969, 5970)	Novel Protein sim. GBank gij103160ipr[S22126 - finger protein unkempt - fruit fly (Drosophila melanogaster)]		UNCLASSIFIED	264910, 55812038, 56181562, 55811957, 264828, 55810784, 264832, 264835, 60432113
2986	86284861 (5971, 5972)				55811957, 264566
2987	86455934 (5973, 5974)			UNCLASSIFIED	264369